DOCUMENT RESUME

ED 152 427 PS 009 850

AUTHOR	Villaume, John; Haney, Walt
TITLE	The Follow Through Planned Variation Experiment; Volume V. Appendix: Analysis of Interim Follow Through Evaluation Reports.
INSTITUTION	Huron Inst., Cambridge, Mass.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C. Office of Planning, Budgeting, and Evaluation.
PUB DATE	Aug 77
CONTRACT	OEC-0-74-0394
NOTE	373p.; Some tables in this document may be marginally legible due to small print size; For related documents, see ED 094 890-891, ED 108 768-769, ED 148 488-493, and PS 009 849
EDRS PRICE	MF-$0.83 HC-$19.41 Plus Postage.
DESCRIPTORS	*Academic Achievement; Bibliographies; *Disadvantaged Youth; *Primary Education; *Program Effectiveness; *Program Evaluation; Summative Evaluation
IDENTIFIERS	*Planned Variation Program; *Project Follow Through

ABSTRACT

This report is the appendix to the last volume in a five volume series on the national evaluation study of the Follow Through Planned Variation Model. The appendix is organized in three sections. Section I provides detailed summaries of five major national Follow Through evaluation reports not published as part of this series. Section II is an annotated bibliography of more than 50 other evaluation reports funded by the United States Office of Education since the inception of the program. Section III contains unannotated citations for other documents concerning Follow Through, including Congressional hearings and reports, items from the Federal Register, reports for the Government Accounting Office, Follow Through evaluation technical papers and feasibility studies, Follow Through conference reports, books and articles, theses, papers presented at meetings and conferences, Follow Through-related materials produced by State Departments of Education or school districts, and Follow Through sponsor monographs. (Author/JMB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

THE FOLLOW THROUGH PLANNED VARIATION EXPERIMENT

Volume V

Appendix Analysis of Interim Follow Through Evaluation Reports

DISCRIMINATION PROHIBITED. No person in the United States shall, on the ground of race, color, or national origin, be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any program or activity receiving Federal Financial Assistance, or be so treated on the basis of sex under most education programs or activities receiving Federal Assistance.

THE FOLLOW THROUGH PLANNED VARIATION EXPERIMENT

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
Joseph A. Califano, *Secretary*
Mary Berry, *Assistant Secretary for Education*
Office of Education
Ernest Boyer, *Commissioner*

THE FOLLOW THROUGH PLANNED VARIATION EXPERIMENT

Volume V Appendix:

Analysis of Interim Follow Through Evaluation Reports

By

John Villaume

Walt Haney

August 1977

Huron Institute
123 Mt. Auburn Street
Cambridge, MA 02138

Prepared for the
Office of Planning, Budgeting and Evaluation
United States Office of Education
Department of Health, Education and Welfare
Contract No. OEC-0-74-0394

5

PREFACE

Over the span of the Follow Through Planned Variation Experiment an enormous amount of information has been produced. Much of it has been used in the preparation of this publication. A series of volumes has been assembled to provide a comprehensive overview of the research and evaluation activities that have taken place.

Volume I, prepared by the U.S. Office of Education, contains a synopsis of the key elements of the longitudinal evaluation and a synthesis of findings. Specificially, Volume I examines the central question posed by the Follow Through research and evaluation: In what ways, to what extent, and at what cost did the various Follow Through approaches to early childhood education affect the learning and behavioral outcomes of disadvantaged children.

Volume II, a four-part set prepared by Abt Associates Inc., contains the complete presentation and treatment of the analysis conducted as part of the national evaluation which began data collection in school years 1970-71 and 1971-72. Part A contains a description of the study, the educational approaches examined, a discussion of the analytic strategies and methods of presenting results, along with a summary of findings. Part B contains a comprehensive discussion of the results. Part C contains additional analyses produced after the work on Parts A and B had been completed. These analyses complement those in Parts A and B and contain the most complete statistical analyses for all cohort streams included in the evaluation. Part C appendix contains many detailed tables of results and supporting material.

Volume III, prepared primarily by the model sponsors in conjunction with the U.S. Office of Education, describes the models and presents, where available, additional evaluative data collected and analyzed by the sponsors.

Volume IV, prepared by RMC, Inc., contains cost data on twenty of the models.

Volume V, prepared by Huron Institute, is a technical history of the national evaulation. Major issues confronting the evaluation and the attempts made to resolve them are examined in some detail. Part V Appendix contains summaries of five earlier Follow Through evaluation reports not published as part of this series and an annotated bibliography of other evaluation reports funded by USOE since the inception of the program.

Table of Contents

Page

I. SUMMARIES OF MAJOR NATIONAL FOLLOW THROUGH EVALUATION REPORTS 4

1.1 John A. Emrick, Philip Sorenson, and Marian S. Stearns. Interim Evaluation of the National Follow Through Program 1969-1971. A Technical Report. Menlo Park, CA: Stanford Research Institute, Feburary, 1973. 7

1.2 Marvin G. Cline, N. Ames, R. Anderson, R. Bales, T. Ferb, M. Joshi, M. Kane, J. Larson, D. Park, E. Proper, L. Stebbins, and C. Stern. Final Report. Education as Experimentation: Evaluation of the Follow Through Planned Variation Model. Volume IA: Early Effects of Follow Through and Volume IB: Monographs. Cambridge, MA: Abt Associates, March 1, 1974. 56

1.3 Marvin G. Cline, N. Ames, R. Anderson, M. Conte, T. Ferb, C. Hatt, M. Joshi, J. Larson, D. Park, E. Proper, L. Stebbins, and C. Stern. Final Report. Education as Experimentation: Evaluation of the Follow Through Planned Variation Model. Volume IIA: Two-Year Effects of Follow Through and Volume IIB: Appendices and Monographs. Cambridge, MA: Abt Associates, May 1, 1975. 139

1.4 Jane A. Stallings, David H. Kaskowitz, and others. Follow Through Classroom Observation Evaluation 1972-73. Menlo Park, CA: Stanford Research Institute, August, 1974. 192

1.5 Robert S. Soar. Final Report. Follow Through Classroom Process Measurement and Pupil Growth (1970-71). Gainesville, FL: College of Education, University of Florida, June, 1973. 251

II. OTHER FOLLOW THROUGH EVALUATION STUDIES FUNDED BY THE OFFICE OF EDUCATION AND THE DEPARTMENT OF HEALTH, EDUCATION AND WELFARE 290

III. OTHER DOCUMENTS 324

3.1 Congressional Hearings and Reports, Items from the Federal Register, and GAO Reports 325

3.2 Follow Through Evaluation Technical Papers and Feasibility Studies 327

3.3 Follow Through-Related Conference Reports 329

Page

3.4 Books and Articles 330

3.5 Theses 336

3.6 Papers Presented at Meetings and Conferences 338

3.7 Follow Through-Related Materials Produced by State Departments of Education or School Districts 344

3.8 Monographs by Follow Through Sponsors or About Particular Follow Through Sponsor Approaches 348

3.9 Miscellanea 354

This appendix is an updated synthesis of two previous bibliographies on the national evaluation of Project Follow Through* The appendix has three sections. The first provides detailed summaries of five major national Follow Through evaluation reports. The second section is comprised of brief descriptions of other FT evaluation studies. Generally the reports described in section II are narrower in scope than those summarized in section I. There are, however, four exceptions to this general pattern. Four documents which we term major reports are described in section II. The discrepancy in the way the nine major reports are handled--five are summarized in detail in section I, while four others are only briefly described in section II--is strictly an artifact of the history of work on this project. After the detailed summaries of the first five reports were prepared, it was decided that briefer descriptions might be more suitable. However, this decision was recently reversed and there has simply not been time available to prepare detailed summaries of the remaining four major reports.

The third section contains unannotated citations for a wide range of other documents concerning Follow Through. These include:

(1) Congressional Hearings and Reports, Items from the Federal Register and Reports from the Government Accounting Office
(2) Follow Through Evaluation Technical Papers and Feasibility Studies
(3) Follow Through Conference Reports
(4) Books and Articles

---

*The two previous draft reports, both by Villaume and Haney are: Appendix to Analysis of Interim Follow Through Evaluation Reports, Cambridge, MA: Huron Institute, August 1976; and Project Follow Through National Evaluation Bibliography, Cambridge, MA: Huron Institute, April 1977.

(5) Theses
(6) Papers Presented at Meetings and Conferences
(7) Follow Through-Related Materials Produced by State Departments of Education or School Districts
(8) Follow Through Sponsor Monographs
(9) Miscellanea

Listings in the second and third sections are not exhaustive. Much material produced on Follow Through has not been published and many Follow Through reports exist in several versions. We have not been able to track down all of this fugitive literature. The entries in this bibliography have been identified (a) in the process of preparing Huron's technical history of the Follow Through evaluation; (b) by a February 1977 computer search of the ERIC system; (c) by a February 1977 computer search of Psychological Abstracts and Dissertation Abstracts; (d) by a search of Huron Institute's Follow Through files; (e) via correspondence and consultation with officials in OE's Office of Planning, Budgeting and Evaluation; and (f) by serendipity.

Several ways in which the bibliography is not exhaustive should be noted. First, the Congressional citations listed fall short of a comprehensive bibliography of Congressional materials on Follow Through. Section 3.1 should therefore not be considered a comprehensive list of documents pertaining to Follow Through's legislative history. The items listed in that section are simply those which we have located in the course of our review of the Follow Through evaluation. Second, the listing of Follow Through sponsor monographs by no means covers all the descriptive materials, Follow Through sponsor-conducted evaluations, and other materials produced by Follow Through sponsors. For example, no effort was made to include references to sponsors' instructional materials.*

---

*The most comprehensive source of information on these sorts of materials is Susan W. Rath, Barbara B. O'Neil, Betty D. Gedney and Josina Osorio, Follow Through: A Resource Guide to Sponsor Models and Materials (Portland, OR: Nero and Associates, Inc., June 1976, 417 pp.). Also see volume III of OE's series, The Follow Through Planned Variation Experiment.

Third, agencies contracted to evaluate Follow Through have produced numerous interim reports, special technical notes, various versions of measuring instruments, and internal reports. This bibliography includes only a sampling of documents of this sort.

The criteria for inclusion in various sections and subsections sometimes overlap. As a result, where an item is placed in the bibliography is sometimes somewhat arbitrary. All citations, however, appear only once.

I. SUMMARIES OF MAJOR NATIONAL FOLLOW THROUGH EVALUATION REPORTS

We have identified nine Follow Through evaluation reports as major reports on the following grounds. First, they have all been sponsored and accepted by the U.S. Office of Education Follow Through Evaluation Office. Second, they are relatively broad in scope. That is, they are based on data collected from a large number of sites affiliated with a relatively large proportion of Follow Through sponsors. Third, for the most part each report's findings are based on data collected over several years. Taken as a whole, these reports are the product of the vast majority of time, energy, and expenditures invested in the national Follow Through evaluation. These nine documents, which we term major national Follow Through evaluation reports are:

John A. Emrick, Phillip Sorenson, and Marian Stearns. Interim Evaluation of the National Follow Through Program 1969-71.

Marvin G. Cline, et al. Education as Experimentation: Evaluation of the Follow Through Planned Variation Model. Volumes IA & IB.

Marvin G. Cline, et al. Education as Experimentation: Evaluation of the Follow Through Planned Variation Model. Volumes IIA & IIB.

Jane A. Stallings, David H. Kaskowitz, et al. Follow Through Classroom Observation Evaluation 1972-73.

Robert S. Soar. Final Report Follow Through Classroom Process Measurement and Pupil Growth (1970-71).

Jane A. Stallings, et al. Follow Through Program Classroom Observation Evaluation 1971-72.

Linda B. Stebbins, et al. Education as Experimentation: A Planned Variation Model. Volumes IIIA & IIIB.

Linda B. Stebbins, et al. Education as Experimentation, Volumes IVA, IVB, IVC, IVD, and the Non-Follow Through Study.

Donald D. Rogers, et al. A Cost Analysis of Follow Through Projects.

12

The first five reports listed are summarized in detail in this section of the appendix. The remaining four are summarized more briefly in section II.

The distinctive manner in which the major Follow Through evaluation reports have been organized and written precludes their being summarized in an identical fashion. Consequently, summaries of them vary somewhat in format. Nevertheless, we have tried to impose some consistency on the format of the summaries by describing each analysis in terms of the following categories.

A. Research Question(s): The substantive issue or question(s) under empirical investigation.

B. Sample: A description of the group of children, schools, communities and/or Follow Through sponsors on which data were collected to invetigate the issues of concern. When possible, a description of how the sample was constructed is included.

C. Variables: A description of the independent and dependent variables under investigation. Where appropriate, summaries also include a description of the covariates or variables which were controlled while the relationship between the independent and dependent variables was examined.

D. Analytical Method: The manner in which the relationship between independent and dependent variables was examined--for example, correlational analysis, analysis of variance (ANOVA), or analysis of covariance (ANCOVA).

E. Comments: Attributes of the study that warrant special attention. Usually this refers to logical or procedural limitations which should, we think, be taken into consideration when reviewing reported results.

Comments on specific studies are limited to those which are peculiar to the analysis or analyses under consideration. Caveats common to Follow Through evaluation studies in general are treated in the technical history of the national evaluation. The technical history takes up issues having to do with the internal and external validity of findings, levels of analyses, the specification and verification of "treatments" (often the independent variables in the analyses summarized), levels of analysis, the adequacy of measurement instruments used to operationalize variables, problems of attrition, and the adequacy of analytical methods. These issues have important implications for the results, and particularly the limitations of virtually all of the analyses summarized here.

F. Results: Summary of the results of the analyses usually as reported or, in some cases, as condensed from reported results.

Although these six descriptive categories have been used whenever possible in section I, the reader should be forewarned that the manner in which the categories have been applied to different "studies" is somewhat arbitrary--for example, if a number of different questions have been addressed using the same sample and set of variables, we have not necessarily summarized them as separate studies. Our general guide for grouping questions, samples, variables, etc., into "studies" was simply to do so in whatever way seemed both compact and easily comprehensible.

1.1 John A. Emrick, Philip Sorenson, and Marian S. Stearns. Interim Evaluation of the National Follow Through Program 1969-1971. A Technical Report. Menlo Park, CA: Stanford Research Institute, Feb., 1973.

This report is the culmination of SRI's work in evaluating the learning outcomes for children participating in Project Follow Through (FT). Numerous SRI reports on FT preceded this one but most were draft reports.

The classroom observation investigation reported here was extended and is reported in Jane Stallings, Follow Through Classroom Observation Evaluation 1972-73, summarized below.

This summary of SRI 1969-71 report is organized into four parts:

I. CHILD-FOCUSED INQUIRIES. . . . . p. 8
II. CLASSROOM OBSERVATION. . . . . p. 27
III. PARENT-FOCUSED INQUIRIES . . . . p. 39
IV. TEACHER-FOCUSED INQUIRIES. . . . . p. 49

I. CHILD-FOCUSED INQUIRIES

A. Research Questions: Questions were formulated at three levels of analysis:

1. National impact of the program: "On the average, is Follow Through as a national program of intervention producing measurable impact?"
2. Sponsor-level impact of the program: "Have at least some of the approaches or planned variations produced a measurable impact?" (p. 75).
3. Site-level impact of the program: "Have some FT projects (particular planned variations employed at particular sites) produced a measurable impact?" (p. 75).

Each of these questions contains several questions concerning one-year or two-year effects. The one-year questions were addressed with data from both Cohort I and Cohort II. Similarly, each was examined with reference to the issue of possible differences between grade streams entering kindergarten or first grade level.

B. Sample:

1. Method of construction: A purposive sample was constructed using the following decision rules in a sequential fashion.
    a. SRI solicited sponsors' judgments concerning sites where respective models were thought to be well implemented. SRI used this as a starting list of sites to be included in the analytical subset.

    b. USOE/FT subsequently designated that some sites should be included and others excluded. The starting list was amended to comply with these recommendations.

    c. SRI adjusted the above amended list:

    -- to obtain at least five sites per sponsor if possible
    -- to maintain a 4:1 distribution of EK to EF sites
    -- to obtain geographic and urban/rural balance
    -- to avoid impractical situations, such as locations where comparisons were not obtained.

    d. USOE/FT subsequently informed SRI that no site with less than one year of operational experience should be included.
2. Description: The sites selected included both Cohort I and Cohort II children.

Cohort I: Sites remaining in the analytical subset after the application of the various decision rules. These sites were distributed as follows:

13 Sponsors:*	01 SS	02 FW	03 UA	05 BC	06 UG	07 UO	08 UK	09 HS	10 FL	11 ED	13 NY	14 SW	15 PI	Totals
Total number of sites/sponsor:	5	3	4	5	1	5	3	3	3	3	2	1	1	39
One- and two-year data:	5	2	3	3	0	3	2	2	2	2	1	2	1	28
Two-year data only:	1	1	1	3	2	3	2	1	1	1	1	0	0	17

Cohort II: Twenty-eight sites remained in the analytic subset distributed as follows:

13 Sponsors:	01 SS	02 FW	03 UA	05 BC	06 UG	07 UO	08 UK	09 HS	10 FL	11 ED	13 NY	14 SW	15 PI	
Sites/Sponsor:	1	3	3	5	0	4	3	2	3	2	2	0	0	28

The number of classrooms included were:

	Cohort I	Cohort II	Total
Entering Kindergarten	337	93	430
Entering First Grade	118	34	152
Total	455	127	582

---

*Code numbers and abbreviations for sponsors are:

01	SS	Self-sponsored
02	FW	Far West Laboratories
03	UA	University of Arizona
05	BC	Bank Street College
06	UG	University of Georgia
07	UO	University of Oregon
08	UK	University of Kansas
09	HS	High/Scope
10	FL	University of Florida
11	ED	Education Development Center
13	NY	New York University
14	SW	Southwest Educational Development Laboratory
15	PI	Parent Implemented

The first and last categories refer to types of sponsors as opposed to the others which are institutional in character.

C. Variables:

1. Independent: FT or NFT experience.

2. Dependent: Eight outcome measures were used.

   "1) Overall achievement--the raw score sum of correct responses to all cognitive test battery items.

   "2) Affect--the scaled sum of seven responses to the attitude inventory (3 = happy, 1 = sad; min. = 7, max. = 21).

   "3) Attendance--the average number of days absent from class in the 1970-1971 school year.

   "4) WRAT--the raw score sum of correct responses to the WRAT" (p. 27).

   "5) Quantitative skills--defined as the raw score sum of correct items pertaining to quantitative concepts--such as numeration, operations (addition, subtraction, etc.) and word problems.

   "6) Reading skills--defined as the raw score sum of correct responses to items requiring reading or reading-related skills (including pre-reading), such skills as alphabet/letter recognition, matching and copying, figure copying, word matching, symbol matching, oddity discrimination.

   "7) Language arts--defined as the raw score sum of correct responses to items requiring language, lexicographic, or grammatical skills such as analogies, word meaning, spelling, and concept activation.

   "8) Cognitive processes--a residual category consisting of the raw score sum of correct responses to items requiring perceptual motor skills and concept identifications" (p. 59).

   "It should be noted that in this organization of instruments into variables, the quantitative, reading, language, and cognitive process variables consist of mutually exclusive subsets of the total achievement variable. The WRAT achievement variable, on the other hand, overlaps with both the total achievement variable and the quantitative, reading, and language variables" (p. 59).

3. Covariates: Outcome measures were adjusted using the following covariates.

   "1) No. of classrooms--the number of classroom level aggregations of pupils for which data are sufficiently complete ($\geq$ 90 percent) for analysis.

"2) Average pupils/classroom--the average number of pupils who met the completeness of data requirement.

"3) Quant. prescore--the raw score (number correct) on the baseline test (maximum varies across cohorts and grade streams).

"4) Cog. Process prescore--the raw cognitive process prescore.

"5) Reading prescore--the raw reading prescore.

"6) Language prescore--the raw language prescore.

"7) Affect prescore--the raw affect prescore.

"8) Age (June '71)--the average age in months of pupils as of June, 1971.

"9) % classroom male--the percentage of boys in the classroom, averaged across classrooms.

"10) % classroom black--the percentage of blacks in the classroom, averaged across classrooms.

"11) % preschool--either (a) the average percentage of pupils who had at least some preschool (CI only), or (b) the average number of months of preschool (CII only).

"12) % parents w/o HS diploma--the percentage of parents of pupils in the class who DO NOT have high school diplomas, averaged across classrooms.

"13) % parents with skilled occupation--the average percentage of parents of pupils in the class who have occupations classified as "skilled", i.e., professional, clerical, manager, sales, craft, etc., but NOT service, laborer, operative, housewife, etc.

"14) % parents black--average percentage of pupil parents who are black.

"15) % poverty eligible--average percentage meeting OEO poverty guidelines as shown in following tabulation.

"16) % head of household employed--the average percentage of heads of households who report they are currently employed (parent interview responses).

"17) % head household male--the average percentage of heads of households who are male" (pp. 82-85).

D. Analytical Method:

1. Overall strategy: FT and NFT comparison groups were known to be non-equivalent. An attempt would have to be made to adjust for these differences. Moreover, different tests had been used at different data collecting points to measure children's academic progress. The analytical method used would also have to be suited for adjusting post-treatment scores on a given test by removing the effects of pre-treatment scores on a different test. The particular method selected to achieve both these purposes was ANCOVA.

2. Model: "Analysis of covariance (ANCOVA) was the method chosen for integrating data from the various sources and performing tests of the major and sub-hypotheses. The specific ANCOVA model employed is a fixed effects, one-way design in which approaches (planned variations) are arranged as the treatment variables, and control variables are used as covariates. Furthermore, within each treatment group in the one-way design, the FT project samples and the corresponding NFT control groups are nested. This analysis design was implemented separately for pupil, parent, and teacher variables and for each cohort sample, grade stream, and assessment interval. Thus, 3 (pupil, parent, teacher) x 2 (Cohort I, II) x 2 (K, EF), or 12 independent analyses were performed on 1971 data, and 2 additional analyses were performed on the 1970 K and EF pupil data. Analyses of teacher and parent variables were not performed on the 1970 Cohort I subset" (p. 76).

E. Comments:

1. Sample: It should be obvious that the analytic subset is the product of administrative dicta, not any well-defined selection process, e.g., random selection. Whether the classrooms involved are representative of the respective cohorts can only be determined empirically. No effort was made--at least none is reported--to determine whether sites, schools, classrooms, or children included in the analytic sample were representative of the sites, schools, classrooms, or children of Cohorts I and II, as a whole.

   In addition to the uncertainty concerning the representativeness of initially included elements vis-a-vis initially excluded elements, the representativeness of the sample may also have been affected by attrition. Students migrated to and from FT classes such that the composition of a classroom included in the analysis did not remain constant. Analyses were performed on only "identic subgroups," i.e., those students for whom 90% of both pre- and posttest data were available. If migration operated randomly, i.e., equivalent types of children moving into and out of different classes, there would be no problem, but comparisons of scores for the "identic" subgroups with those of the total classroom population indicate that migration was not a random phenomenon.

In particular, while identic subgroups proved to be no different than total classroom populations in terms of baseline scores on the six achievement outcome measures, their performance was significantly (.05 or better) different on posttest outcome measures for Cohort I-EK, Cohort II-EK, and Cohort II-EF. There were noticeable differences for Cohort I-EF as well. In every case, the direction of differences indicates that the identic subgroups performed better on the achievement measures. SRI suggests that "One interpretation of this phenomenon is that the slowest pupils are most likely to be subject to attrition and/or pupils transferring into the program are academically behind the continuing (identic) pupils" (p. A-10).

On two accounts, then, the representativeness of the analytical subset is called into question. And, as long as it remains in question, both the internal and external validity of results are also open to question.

2. Variables: The appropriateness of dependent variables and covariates to be used is a question that has beleaguered FT evaluation throughout its history. The pertinent points of the controversy surrounding that question have been dealt with in the text. We only mention here that (1) the set of child outcome measures used deals with an extremely circumscribed number of potential and/or intended FT program outcomes, and that (2) the appropriateness of the omnibus set of covariates employed is a largely undiscussed issue.

Although several outcome measures are given labels which imply an underlying conceptual dimension, e.g., "quantitative," "reading," "language" skills, and "cognitive processes," these labels are strictly speaking only indicators of particular questions used to construct respective scores. The underlying conceptual dimensions for each are never clearly specified; the degree to which the underlying conceptual dimension is validly measured consequently can not be determined.

The label "cognitive processes" is a good example of the operational nature of the concepts referred to. SRI's report defines this variable as "a residual category consisting of the raw score sum of correct responses to items requiring perceptual motor skills and concept identifications" (p. 59). Although it is not clear how exactly any of the skill-labeled variables were constructed, it appears from the "residual" used in the definition that this variable was constructed by adding the correct responses to all questions that were not used to construct the "quantitative skills," "reading skills," and "language arts" outcome variables. This is a clean operational definition, but the concept reflected by this residual measure is certainly not immediately apparent. Comparisons of FT with NFT groups on this measure show that children of the respective groups differ or do not differ on some set of scores, the meaning of which is not clear. Interpretations based on such comparisons are not likely to be particularly meaningful in the fundamental sense of that term.

What has been said about "cognitive processes" can also be said about most of the other outcome measures--the four skill-labeled variables and "affect." The same issues could be raised concerning many of the covariates used.

To return to the issue of reliability, SRI describes the reliability of the six achievement outcome measures in terms of coefficient alpha. Alpha reliabilities are quite acceptable for five of the six outcome measures across the two cohorts and grade structures within cohorts. The alpha reliability of the cognitive processes' measure, which ranges from a low of .580 to a high of .760, is the lowest of the six. No other forms of reliability, for example, test-retest, are reported.

The reliability of attendance and affect are not discussed. Nor is the reliability of any of the covariates treated.

3. Analytical method:

SRI makes overall FT/NFT comparisons in two ways. Both methods need to be made explicit in order to make the results understandable. The first method of overall, i.e., national FT/NFT comparisons, calculates "overall cohort averages (computed by summing across projects)" for each of the eight outcome variables. The calculated values represent adjusted overall FT/NFT differences. Positive values indicate that FT classrooms have higher scores on a given outcome measure and thus constitute FT-favoring differences. Negative values indicate NFT-favoring differences. (The opposite applies for the absence outcome where a negative difference is preferable.)

The report is not at all clear as to exactly how cohort averages were summed across sponsors and/or sites. Are the overall adjusted differences simply the weighted mean of the site level differences? Or was the overall adjusted difference calculated by averaging classroom differences across all sites? Or was some other "summing" process used? Moreover, it remains unclear whether adjustments are made by using all covariates in calculating the differences on each outcome measure or whether selected covariates are used for different outcome measures. This same ambiguity exists in all three levels of analysis. The lack of answers to these questions makes it difficult to interpret the overall adjusted differences that are reported.

The second method used to make overall FT/NFT comparisons is to express the number of site level, FT-favoring differences as a percentage of all the sites involved in a particular cohort analysis. Thus, if 15 of 24 site level differences on WRAT scores favor FT, the percentage 54% is reported for this variable. Percentages are reported for each of the eight outcome variables. The arbitrariness of this counting procedure can be demonstrated by considering an alternative counting method.

Instead of noting the sign of differences, statistically significant differences might have been counted. Percentages of sites reflecting a statistical difference--both FT- and NFT-favoring--could then be reported. To take a particular case, SRI reports that 54% of the Cohort I-EK sites had FT-favoring differences on the WRAT outcome after two years. The number of statistically significant, FT-favoring differences on this outcome is 2. As a percentage, 2 of 28, or 7%, of the sites involved in this analysis record FT-favoring differences. Conversely, 4 sites have statistically significant NFT-favoring differences. Thus, 4 of 28, or 14%, of the site differences are NFT-favoring. The "results" of this alternative counting and reporting method are dramatically different from those used by SRI.

Percentage of Site-Level
Differences for the Two-Year WRAT Outcomes

	FT-Favoring	NFT-Favoring
SRI method based on sign of differences	54%	46%
Alternative method based on statistical significance	7%	14%

The two methods lead to considerably different interpretations.

SRI contends that the extremely low power of statistical analysis (along with other factors) makes it "preferable to interpret these data . . . in the context of the overall, as opposed to 'significantly only' results. . . ." (p. 293). The implication is that statistical significance may not be an appropriate basis for counting differences. While we do not contend with SRI's observation, we do point out that "negative impacts" may likewise remain undetected. If real differences remain undetected due to the lack of power in the significance tests, they may be negative, i.e., NFT-favoring, as well as positive or FT-favoring. It is more important to note that SRI's observation is not the central issue. It should not obscure the fact that the counting method employed is highly arbitrary and that an alternative (and perhaps more conventional) counting method yields dramatically different results. The problems just discussed are related to two more fundamental problems encountered in dealing with overall FT/NFT comparisons as they are reported.

The first of these more fundamental problems is the problem of how to interpret the magnitude of differences reported--in themselves and with reference to one another. Reported differences are adjusted, raw score differences. They thus reflect the differences of scale underlying the measurement of each of the eight outcome variables. Certain questions arise. What does an overall FT/NFT difference of -0.1 on reading skills mean? Can this difference be interpreted in light of a +0.8 difference on overall achievement? How do these compare with a - 0.1 difference on affect? There are no readily apparent answers to these questions.

The second, more fundamental problem stems from treating unequivalent outcomes as equivalent. When confronted by eight overall difference scores for each of the cohorts, with their respective grade streams, with both one- and two-year effects in some cases, the mind seeks a strategy for organizing the data in order to deal with the many comparisons that can be made. There is an almost inherent tendency to treat the eight outcome differences as equivalent. It is, therefore, easy to understand the report's doing so. However, the appropriateness of treating outcomes of an obviously different nature as equivalent needs to be considered. "Box-scoring" cohort performances in this manner implicitly assumes, for example, that the relative importance of affect is the same as that of quantitative skills, or that of each of the other outcomes for that matter. Such an assumption would certainly not receive universal assent. It is, or should be, more a matter of explicit judgment than one of implicit assumption. Moreover, the eight outcome measures are not mutually exclusive. For example, overall achievement obviously includes WRAT, quantitative, cognitive processes, reading skills, and language arts. Thus, double counting occurs when a box score statement like "3 of 8 outcome differences favor FT" is made.

Both because of the judgments involved and because of the lack of mutual exclusiveness among the various outcome measures, the strategy of treating them all as equivalent is highly questionable. Box-score-based comparisons which rely on treating them in such a way are similarly questionable.

We have dwelt on the above matters in order to alert the reader to the fact that procedures used to report overall FT/NFT comparisons involve summing or assimilating principles. These principles need to be made specific and their consequences noted. The consequences we have identified suggest that the reported results of the overall FT/NFT comparisons must be carefully considered. The reader will have to make his own judgment whether they warrant interpretation. If he decides they do, we hope that he enters the realm of interpretation with a very keen awareness of the fragile structure of assumptions underlying any interpretation.

Most of what has been said concerning overall FT/NFT comparisons and the way in which these comparisons are interpreted can also be said for comparisons at the sponsor and site levels.

F. Results: Results are reported with reference to the three levels of analysis incorporated in the research questions: national, sponsor, and site level.

1. Results at the national or overall level (pp. 275-293):

   a. Cohort I-EK

      (1) Two-year effects (Fall 1969 to Spring 1971). SRI sums the 28-site level, FT/NFT differences on each of the eight outcome measures. "These results are mixed, with little evidence of FT impact on child outcome variables" (p. 280). Although differences favor FT classrooms on achievement, attendance, quantitative skills, and cognitive processes, "all of these differences are small and not especially noteworthy" (p. 280).

      (2) One-year effects (Fall 1969 to Spring 1970. "FT-favoring outcomes are noted in 75 percent (9) of the projects for the quantitative measure and in 83 percent (9) of the projects for the cognitive processes variable. However, the average FT-favoring difference per project for these variables is relatively small. The results for the remaining variables are not particularly noteworthy" (p. 281).

      (3) Comparison of two-year versus one-year effects. A subset of the sites (12 of 28) involved in the analysis of two-year effects were also involved in the analysis

of one-year effects. When both the magnitude of overall differences between FT and NFT children and the percentage of sites at which differences favor FT children are considered, it appears that, ". . . with the exceptions of the cognitive processes variable and the language variable, FT children show a greater advantage over their NFT counterparts after two years than they do after one year" (p. 281).

b. Cohort I-EF

(1) Two-year effects (Fall 1969 to Spring 1971). "The only overall outcomes in favor of FT were those for the attendance. . . ." (p. 283). SRI reports an FT-favoring difference for quantitative skills, but this is not consistent with the data they present.

(2) One-year effects (Fall 1969 to Spring 1970). "The pattern of first-year results is essentially the same as that noted for the second-year data. Most of the variables display NFT-favoring trends. . . ." (p. 283). Only attendance and affect differences are FT-favoring.

(3) Comparison of two-year versus one-year effects. SRI points out that only seven sites were involved in the analysis of both one- and two-year effects. The confidence one can place in the results is thus quite limited. However, SRI comments ". . . second-year effects for Cohort I-K were more favorable than first-year effects, suggesting a cumulative positive impact for FT. [But] the opposite is true for the present Cohort I-EF projects, where there is evidence of a positive decrement; second-year outcomes more frequently reflect NFT-favoring trends. . . ." (p. 283).

c. Cohort II. Note that Cohort II children finished only one year of FT experience by Spring 1971--the last data collection time point for this analysis. Thus, only one-year effects can be reported for this cohort.

(1) Entering Kindergarten (Fall 1970). Averaged overall FT/NFT differences favor FT children on most of the outcomes. The only difference suggesting a preferred performance by the NFT children is affect scores.

(2) Entering First Grade (Fall 1970). Only four sites fall into this category. "Summary statistics are likely to be unreliable." SRI does report that "with the exception of the attendance variable, all of the child outcome variables show FT-favoring differences" (p. 283).

2. Results at the sponsor level:

Results of analysis at this level are the product of aggregating observations so as to compare FT/NFT differences within each sponsor. The method of aggregation is not specified. SRI does not indicate whether (1) children, classroom, or site scores were pooled and FT/NFT differences then calculated, or (2) differences at each of these three levels were somehow aggregated. This ambiguity renders the actual differences difficult to interpret.

Potential analyses can be summarized as follows:

Grade Stream	Cohort I (Effects)		Cohort II (Effects)
EK	1 yr.	2 yr.	1 yr.
EF	1 yr.	2 yr.	1 yr.

Summary sponsor results are presented separately for each of these potential six analyses. However, few sponsors had enough sites (at least two for a particular cohort and grade stream) to permit all possible analyses. Thus, summarized sponsor results include only some of these analyses.

Treatment effects are reported in terms of adjusted FT/NFT differences on each of the eight outcomes. Statistical tests were performed to determine whether these differences were significantly different from 0.0 at the .05 level of confidence.

In general, very few significant results were obtained--at least very few in terms of the total number of comparisons involved. All significant differences were obtained with Cohort I-EK children at the end of two years. These were:

-- Bank Street College--quantitative skills (FT);

-- High/Scope--attendance (FT); WRAT total (NFT); reading skills (NFT); language arts (NFT);

-- New York University--quantitative skills (FT).

As the abbreviations in parentheses indicate, all significant differences were not FT-favoring.

a. Far West Laboratory for Educational Research and Development (pp. 108-111):

(1) Cohort I, EK, two-year effects. All adjusted differences fail to be significantly different (.05 level) from 0.0 indicating no FT impact.

(2) Cohort II, EK, one-year effects. All adjusted differences fail to be significantly different (.05 level) from 0.0 indicating no FT impact.

"The results summarized . . . indicate that evidence at the sponsor level is not particularly favorable to this model. In particular, no pupil level results are indicated in either Cohort I or Cohort II analyses."

b. University of Arizona: Tucson Early Education Model (pp. 130-132):

(1) Cohort I, EK, two-year effects. All adjusted differences proved to be less than significantly different from 0.0, indicating no FT impact.

(2) Cohort I, EF, two-year effects. All adjusted differences were not significantly different from 0.0 indicating no FT impact.

"These results show that the University of Arizona model has not produced identifiable impact on pupil outcome after two years."

c. Bank Street College of Education Approach (pp. 148-149):

Cohort I, EK, two-year effects. The only outcome difference significantly different from 0.0 was cognitive processes. In this case, the difference was positive indicating a desired FT impact.

d. University of Georgia, Mathemagenic Activities Program (pp. 153-155):

Since there was only one site from which data were collected for this analysis, the ". . . risk of faulty interpretations is considered very high." Analysis based on this single site was for Cohort I, EF, two-year effects: No statistically significant differences were noted.

e. University of Oregon, Engelmann/Becker Model for Direct Instruction (pp. 171-175):

SRI comments:

"It is unfortunate that the FT and NFT samples for the Cohort I and Cohort II projects in this evaluation were so badly matched."

* * * * *

"As discussed in the interpretation of project data, these FT and NFT samples represented distinct population subgroups. Hence, the results of significance tests on pupil outcomes are likely invalid. . . . That is, the FT samples were characterized by Black children from poor families with low baseline scores, whereas the NFT groups were characterized by non-Black children with higher baseline scores from families that can scarcely be characterized as disadvantaged" (p. 172).

(1) Cohort I, EK, two-year effects. No significant differences were observed on any of the eight outcome measures.

(2) Cohort I, EF, two-year effects. No significant differences were observed on any of the eight outcome measures.

f. University of Kansas, Behavioral Analysis Approach (pp. 186-187):

Cohort I, EK, two-year effects. A significant difference was found in the cognitive processes outcome. All other differences were insignificant.

SRI comments:

"Because of missing data and the lack of comparability of the comparison groups, these results are probably invalid. Thus, we cannot confidently make statements regarding the relative impacts of the model at this time."

g. High/Scope Educational Research Foundation, Cognitively Oriented Curriculum Model (pp. 198-200):

"Since the serious comparison group and data collection problems noted or suspected for two of the three samples lead to our recommendation that these samples not be included as evaluative evidence regarding the model's effects to date, our judgments and comments must be based on analysis of a single project" (p. 199).

Cohort I, EF, two-year effects. Four significant differences were detected at this single site. One--attendance--favored FT. The other three--WRAT total, reading skills, and language arts--favored NFT.

"It is possible that the current evaluation variables are not appropriately sensitive to many of the cognitive behaviors that the model purports to develop. Also, since process factors showed high FT/NFT similarity, it may be that the model was not well implemented in the FT school. Nonetheless, the data show FT pupils are clearly behind NFT in the learning of certain basic skills" (p. 200).

h. University of Florida, Florida Parent Education Model (pp. 215-217):

Cohort I, EK, two-year effects. No significant differences were observed, indicating no identifiable FT impact after two years.

i. Education Development Center, EDC Open Education Program (pp. 233-235):

Cohort I, EK, two-year effects. No significant differences were obtained, suggesting no identifiable FT impact after two years.

j. New York University, Interdependent Learning Model (pp. 243-246):

Cohort I, EK, two-year effects. A significant difference in the quantitative skills was observed. This positive difference indicates an FT impact in this area.

k. Southwest Educational Development Laboratory (pp. 250-251):

Cohort I, EK, two-year effects. FT-favoring differences were obtained for four child-outcome measures: overall achievement, total WRAT score, quantitative skills, and reading skills. SRI comments, however, that since all the data are taken ". . . from a single sample within a single project, we feel no conclusions can be justified at this time" (p. 251).

3. Results at the site level:

The focus of SRI's evaluation is at the site level. The major portion of the results chapter is given to site level analysis. Each site is described in terms of its size, racial composition, location with reference to the nearest Standard Metropolitan Statistical Area, average annual pupil expenditure, Parent Advisory Council membership per pupil ratio, etc. Considerable attention is paid to the quality of FT/NFT matching. In sum, the report provides the kind of detail necessary to sensibly interpret findings concerning FT/NFT differences on the eight child outcome measures.

Summarizing the detailed information provided poses a dilemma. On the one hand, something less than a reproduction of the 175 pages given to the discussion of site level findings is required. On the other hand, precisely because the detailed information is so useful and necessary for interpreting reported findings, any effort to reduce it to synoptic form is bound to separate the findings from the contextual information needed to interpret them. There is no satisfactory solution to this dilemma.

In providing our summary, we have resorted to the following procedure. We extract from SRI's tables data on outcome measures for each site level. The results are grouped according to cohort, grade stream, and one- or two-year effects. The tables also include information on the quality of matching of FT and NFT classrooms in the form of ratings as "GOOD," "MOD" (moderate), and "POOR." A word of explanation concerning these ratings is necessary.

The quality of matching ratings were based on inspection of seven baseline demographic variables:

-- percentage of students with preschool experience,
-- percentage of parents without high school diplomas,
-- percentage of parents in skilled occupations,
-- percentage of Black parents,
-- percentage of parents who are poverty eligible,
-- percentage of heads of household currently employed, and
-- percentage of heads of household who are male.

Discrepancies of 10% or more between FT and NFT classrooms were then noted. If a FT classroom was different from its NFT comparison classroom on three or fewer of the seven baseline demographic variables, the FT/NFT match was considered "GOOD." If the two classrooms differed in four to five variables, the match was rated "MOD." The match was considered "POOR" if there were six or seven discrepancies of over 10%.

While this information helps somewhat in interpreting outcome differences, it ignores the sources of the mismatch. Moreover, the summary tables we provide likewise ignore mismatches that are the result of children's different entering test scores and other classroom characteristics. We emphasize our discontent with how we have summarized site level findings in this regard and refer the reader who is interested to the detailed information provided in the report.

One additional word of explanation concerning the following table is necessary. The left most column contains "project codes." These stand for:

SS--self-sponsored project
FW--Far West Laboratories
UA--University of Arizona
BC--Bank Street College
UO--University of Oregon
UK--University of Kansas
PI--Parent Implemented
HS--High/Scope
UF--University of Florida
ED--Education Development Center
NY--New York University
SW--Southwest Educational Development Laboratory

(a), (b), (c), and (d) refer to particular sites of a given sponsor.

SUMMARY OF INTERIM EFFECTS FOR COHORT I, KINDERGARTEN*

PROJECT CODE	MATCH	CHILD MEASURES ACHIEVE-MENT	AFFECT	ATTEN-DANCE	WRAT	QUANTI-TATIVE	COG. PROCESS	READING	LAN-GUAGE
TWO YEAR OUTCOMES									
SS(b)†	GOOD	8.0	-.1	-12.0°	3.4	3.7	.6	3.1	.7
SS(c)†	MOD	14.3°	.5	-3.8	6.6°	2.4	.1	8.8°	2.9
SS(d)†	GOOD	-11.1	-.5	-1.9	-7.5°	-3.2	-.4	-6.7°	-.5
SS(e)†	POOR	-10.7	-.6	7.4°	-6.1	-1.9	-.4	-4.9	-3.4
FW(a)	GOOD	-4.1	.7	5.7°	-7.4	.1	-.6	-2.3	-1.6
FW(b)†	GOOD	5.5	.8	-1.7	1.6	.9	.3	2.0	2.3
FW(c)	MOD	9.8	1.2	-.3	3.9	6.5°	.1	0	3.3
UA(a)	GOOD	-2.1	-.1	-3.9	-3.3	-1.8	.6	-1.7	.4
UA(d)†	MOD	-1.2	-1.1	2.1	-1.2	3.4	.7	-2.5	-2.8
BC(a)	POOR	-17.4	.3	1.9	-12.5	-7.0	1.7°	-8.4	-3.3
BC(b)	MOD	-12.9	-2.6°	-1.3	-8.4°	-4.2	.4	-7.6°	-1.6
BC(c)	MOD	8.8	1.0	1.9	2.2	9.3°	-.2	.5	-.7
BC(e)	MOD	2.3	-2.0°	.5	.5	1.2	-.2	0	1.4
UO(a)	POOR	5.1	1.1	-10.5°	5.9	-.2	-.5	4.8	1.0
UO(b)	MOD	-5.4	-1.1	3.1	-1.1	2.5	-.8	-3.7	-3.4
UO(c)	GOOD	8.4	-1.4	-2.2	6.0	-.4	.3	5.6	2.8
UK(a)	MOD	-10.1	.1	-1.1	-8.2	1.6	.3	-6.8	-5.0
UK(b)†	POOR	8.6	.5	1.6	5.2	4.3	1.0	2.6	.7
UK(c)†	GOOD	-1.7	-.7	-4.6	-1.3	1.3	.7	-.7	-3.0
HS(b)	MOD	-27.8°	1.9°	-8.0°	-19.2°	-8.2°	.7	-12.5°	-7.3°
UF(a)	MOD	-16.0°	.2	-15.2°	-9.4°	-3.5	-.1	-7.4	-4.7°
UF(c)†	GOOD	7.2	.4	-1.2	4.9	.6	.5	4.7	1.3
ED(b)†	GOOD	-.8	-.9	1.4	-2.5	2.1	.6	-2.7	-.8
ED(c)†	GOOD	21.0	.7	1.7	9.7	8.6°	.9	10.3	1.5
NY(a)	GOOD	8.1	.3	-7.5	2.6	5.2°	-.5	2.6	.8
NY(b)	GOOD	-.7	-.4	5.5°	.6	-.3	.2	1.5	-2.3
SW	MOD	24.9°	-1.7	-5.0	14.8°	13.3°	-.2	9.8°	2.0
PI†	MOD	12.7	1.5	-2.7	8.2	1.0	.1	8.6	2.8
ONE-YEAR OUTCOMES									
SS(b)	GOOD	11.6°	2.9°	-.7	4.7	3.3°	.5	5.6	2.0°
SS(c)	MOD	-2.2	-2.5°	3.3	-.4	-.9	.2	-1.0	-.1
SS(d)	GOOD	5.2	-.7	-7.5	1.8	1.9	.4	2.7	-.6
SS(e)	POOR	3.0	1.7	-1.7	2.5	.2	.1	3.2	-.6
FW(b)	GOOD	-1.4	1.4	-2.9	-1.2	1.2	0	-3.6	.6
UA(d)	MOD	3.3	.7	-2.7	1.2	.9	.5	1.0	.1
UK(b)	POOR	5.4	.9	9.8°	-.6	4.1	.7	-.5	1.1
UK(c)	GOOD	15.9°	-1.3	2.0	10.1°	3.5°	1.0	10.0°	1.4
UF(c)	GOOD	-3.9	-.4	3.8	-4.6	1.5	.4	-5.6	.3
ED(b)	GOOD	-4.9	-1.2	11.2°	-.4	-.3	.2	-4.7	-.5
ED(c)	MOD	7.8	-1.4	-4.9	1.7	2.0	.3	5.9	.2
PI	MOD	-5.8	-2.3	-1.2	-.6	-3.6	0	-1.9	.2

°Difference reaching significance--i.e., $p < .05$.

†Projects for which one-year outcomes are also given.

---

*Taken from Table 96, p. 276.

SUMMARY OF INTERIM EFFECTS FOR COHORT I,

ENTERING FIRST*

PROJECT CODE	CHILD MEASURES								
	MATCH	ACHIEVEMENT	AFFECT	ATTENDANCE	WRAT	QUANTITATIVE	COG. PROCESS	READING	LANGUAGE
TWO-YEAR OUTCOMES									
SS(a)†	MOD	9.7	.8	-.4	4.4	7.2*		1.8	.6
UA(b)	MOD	16.6	.8	-3.2	6.0	1.4		12.7	2.6
UA(c)†	MOD	-15.2	-.3	.1	-3.9	-2.7		-13.7	1.2
BC(d)†	GOOD	20.1	-.3	-3.0	2.7	6.0*		9.7	2.0
UG	MOD	-2.8	.6	2.1	-1.6	-3.5		1.0	-.3
UO(d)†	POOR	6.5	-.5	-.9	2.3	-4.2		11.6	-.4
UO(e)	--	-7.0	-2.2*	2.8	-2.9	-1.2		-2.6	-3.4
NS(a)†	MOD	-20.1	-1.1	-11.2*	-11.2*	3.6		-20.0*	-3.3
NS(b)	MOD	-14.1	-8.2*	-3.7	-6.2	-5.5		-7.3	-1.3
UF(b)	GOOD	11.3	-.7	-6.5	4.2	2.8		6.5	3.1
ED(a)†	MOD	-5.5	-.3	-2.3	-3.4	-.6		-4.5	-.4
ONE-YEAR OUTCOMES									
SS(a)	MOD	-3.8	.5	3.4	-3.9	-1.6	-.3	-2.9	1.2
UA(c)	MOD	-9.9*	3.3*	2.8	-3.2	-3.8	-1.4*	-2.1	-2.9*
BC(d)	GOOD	.8	1.7	-5.8	.5	.4	-.4	.4	.6
UO(d)	POOR	10.6*	-.3	-9.3*	4.1	2.4	1.1	5.3	1.9
NS(a)	MOD	-6.9	3.8*	-1.6	-5.2	-1.0	.5	-5.5	-1.1
NS(b)	MOD	-4.3	-1.8	-2.4	-1.9	-2.3	-.6	-.3	-1.3
ED(a)	MOD	4.8	-.5	.1	1.2	.3	.9*	3.7	-.1

---

*$p < .05$.

† Projects for which one-year outcomes are also given.

---

*Taken from Table 98, p. 278.

SUMMARY OF INTERIM EFFECTS FOR COHORT II, KINDERGARTEN

(ONE-YEAR OUTCOMES) *

		CHILD MEASURES							
PROJECT CODE	MATCH	ACHIEVE-MENT	AFFECT	ATTEN-DANCE	WRAT	QUANTI-TATIVE	COG. PROCESS	READING	LAN-GUAGE
FW(a)	GOOD	.5	-.9	7.8*	4.6	1.1	-1.1*	1.9	-.6
FW(b)	GOOD	-5.1	-1.2	-2.9	-4.2	-1.8	0	-3.2	-.3
FW(c)	MOD	10.4	-.6	1.9	-1.8	3.0	2.6*	3.1	1.9
BC(c)	GOOD	9.5	.2	6.0	7.6	4.4	1.0	3.7	.8
UO(a)	MOD	27.4	1.9	-7.5	23.1	7.8	.1	16.7	3.6
UF(a)	GOOD	31.4*	1.4	-1.4	20.6*	7.1*	.8	20.6*	3.0*
ED(b)	GOOD	-13.1	-2.3	-7.5	-7.1	-5.9*	-.8	-4.9	-1.6
NY(a)	MOD	8.5	.3	-5.2	3.8	2.8	.6	3.0	2.1

---

*p < .05.

SUMMARY OF INTERIM EFFECTS FOR COHORT II, ENTERING FIRST

(ONE-YEAR OUTCOMES) **

		CHILD MEASURES							
PROJECT CODE	MATCH	ACHIEVE-MENT	AFFECT	ATTEN-DANCE	WRAT	QUANTI-TATIVE	COG. PROCESS	READING	LAN-GUAGE
UA(c)	GOOD	-5.6	2.1*	-5.0	-1.2	-1.6	-1.1	-2.9	-1.0
BC(d)	GOOD	21.5*	1.1	4.1	15.1*	7.8	.5	13.2*	4.3*
UO(d)	MOD	52.3	1.9	-.4	42.9*	22.1	.5	29.6	11.7
UF(b)	POOR	25.9*	2.0	2.4	17.0*	11.9*	.7	13.3*	5.3*

---

*p < .05.

---

*Taken from Table 97, p. 277.

**Taken from Table 99, p. 279.

II. CLASSROOM OBSERVATION (especially pp. 43-45, 52-53, 297-308)

A. Research Questions:

1. Have respective sponsor models been consistently implemented by each of the sponsors? Are classroom activities at different sites of a given sponsor similar to classroom activities at the sponsors' other sites?

2. Are the sponsors' various implemented models discernibly different from one another? Do classroom activities of the different sponsors differ appreciably from one another?

3. Are the classroom activities of the respective sponsors discernibly different from activities in NFT classrooms?

B. Sample:

1. Method of construction: Detail provided in the report is less than sufficient to determine how the sample described below was constructed.

2. Description: The sample involved in the implementation investigation is not the same as that used for analyzing outcomes. It is a much smaller sample composed of 123 classrooms--60 from Cohort I and 63 from Cohort II. Observations were undertaken at 17 sites. Classrooms of 9 different sponsors were observed--classrooms in at least two sites were observed for 7 of these 9 sponsors. Some of the observed classrooms were not included in the outcome analysis sample. SRI suggests that while the implementation sample was not fully tested within the outcome analysis sample, "overlap is substantial." Detailed information in this regard is not provided.

C. Variables:

Five classroom process variables were constructed. Trained observers recorded classroom activities during eight five-minute intervals in each classroom. These observations were made using a Classroom Observation Instrument which permitted recording who (persons in the classroom--teachers, aides, and children) did what (e.g., inform, help, cooperate, refuse, etc.) with whom (independently or with others) and how (e.g., negative, angry, guide to alternative, etc.). Forty-one observational statements were identified as important descriptions of classroom activity. Scores on these 41 descriptions became "classroom process variables." Classroom scores on each of these 41 classroom process variables were intercorrelated. The correlation matrix was subjected to principal, component factor analysis. A varimax rotation on five factors accounted for 57% of the matrix variance. The five identified factors were:

(1) self-regulatory, (2) child-initiated interactions, (3) programmed/academic, (4) expressive, and (5) child self-learning. Scores on these factors were used for examining whether activities in sponsors' classrooms were alike, distinguishable from those in other sponsors' classrooms, and discernibly different from NFT classroom activities.

D. Analytical Method: Different analytical procedures were followed in examining each of the research questions.

1. Consistency of given sponsors' implementation across sites: "Visual examination of the factor profiles," a "judgmental interpretation" of factor scores, was used.

2. Discernible differences among sponsors:

   a. Classroom scores on the self-regulatory and the programmed/academic factors were pooled for each sponsor. Scores for each sponsor were expressed in standard deviations and graphed for visual inspection.

   b. Sponsor means on each of the classroom process variables were tested for differences using analysis of variance.

   c. Sponsor means on each of the 41 classroom process variables were further subjected to the Newman-Kuels method of multiple range testing to identify differences among sponsors.

3. Discernible differences between FT and NFT classroom activities:

   a. FT and NFT classroom scores on the self-regulatory and programmed/academic factors were pooled, expressed in standard deviation units, and graphed for visual inspection.

   b. In addition, analysis of variance was performed on each of the 41 classroom process variables to test the difference between average FT versus average NFT classroom scores. Classroom process variables on which the average of the 97 FT classrooms differed significantly ($p < .05$) from the average of the 26 NFT classrooms were identified and logically ordered according to the degree to which they were probably related to the number of adults in a classroom.

   c. Sponsor means and the mean of NFT classrooms were also rank-ordered. Respective sponsor means could thus be compared with the NFT mean.

E. Comments:

1. Sample: Since the method in which the sample was constructed is not specified, it is impossible to say whether results obtained are generalizable to the larger populations of all FT/NFT classrooms or all classrooms of a given sponsor.

2. Variables: Since all the analyses rest upon the five factors identified, some general comments concerning these factors are in order.

   First, although all the details concerning the factor analysis are not provided, the factor labels represent constructs which seem face-valid in light of the positive and negative loadings of the classroom process variables on each factor.

   Second, using factor analysis to identify the dimensions upon which to base FT/NFT and sponsor-by-sponsor comparisons, as well as the determination of the degree of within-sponsor consistency of implementation, is at odds with the concept of planned variation. Planned variation implies that models are defined. What goes on in a classroom should be specified by the defined model, not by factor analysis.*

   Third, we realize that models were not that well defined early on. Statements of what each sponsor intended to accomplish in the classroom and operational statements concerning the classroom activities to be engaged in order to achieve objectives were not available. Consequently, we appreciate the need to use the back door, so to speak, to get at implementation questions. However, it should not be forgotten that a back door approach was used.

3. Analytical Method:

   As conducted, the analysis can at best indicate the degree of consistency of implementation, differentiation among sponsors, and differentiation between FT/NFT. That is a lot, but it is considerably less than answering the fundamental process questions of whether sponsor models are well-implemented and, if so, how outcomes are related to those models. In order to answer those questions, the link between particular classroom activities and particular outcomes would need to be examined. No such examination is undertaken in SRI's report. As the report states, "in the current situation when we encounter program failures or negative results, we have no way of knowing, or sometimes even guessing, why" (p. 315). The same might be said for program successes or positive results as well.

   The following comments are organized according to the research questions that were addressed:

---

*We acknolwedge the one or two exceptions to this statement--i.e., those models which emphasize working with parents as opposed to structuring classroom activities.

a. Consistency of sponsor implementation. Results are reported for seven of the nine sponsors involved. Results were not reported for Education Development Center and New York University because the observation sample included only one site for each of these two sponsors. The evidence presented by SRI in this regard is a summary of "the results of judgmental interpretation [of] factor scores across classrooms within and between projects . . ." (p. 299). Factor scores themselves are not reported. Consequently, how apparent the "apparently large" or "apparently small inter-project differences" are cannot be determined.

b. Discernible differences among sponsors. Since within-sponsor consistency was found on only two of the five factors, self-regulatory and programmed/academic, graphs were understandably prepared and used for comparing sponsors on only these two factors. SRI suggests that the graphs demonstrate inter-sponsor differences. Certainly some sponsors appear to differ from most of the others. However, it is not clear that UO, HS, and UF differ from one another or that UA, ED, and NY are different on the self-regulatory factor. It might more easily be argued that as groups of three these two groups differ from one another. Similarly, it is not clear from our inspection of the graph that BC, UF, ED, and NY are all that different from one another or that FW, VA, and HS are different on the programmed/academic factor. Again, as groups these sets of sponsors do appear to be different.

Perhaps the most important comment we can make concerning SRI's particular use of a bar graph to demonstrate inter-sponsor comparisons is that, as an analytical method, visual inspection can be susceptible to various, if not conflicting, interpretations.

SRI reports that analysis of variance was used to test the differences among sponsor means on each of the 41 classroom process variables. The results of this analysis, however, are not reported.

SRI does report the results of a multiple range test of sponsor means. These results are perhaps the most informative concerning inter-sponsor difference.

c. Discernible differences between FT and NFT classroom activities. The question of whether or not FT classroom activities differ significantly from NFT classroom activities can be examined several ways. SRI's method is to pool all FT classroom scores on given factors and compare this pooled score with a pooled NFT score on the same factors. The pooled score contrast unfortunately fails to allow comparisons of the degree to which various sponsors' classroom activities do or do not differ from what goes on in an ordinary classroom. In terms of planned variation, that is the paramount question. Whether FT classrooms taken as a whole differ from NFT classrooms may be interesting information, but it does not help identify significant treatments or treatment components that produce desired outcomes. Consequently, the bar graphs displaying the differences between FT and NFT classrooms on the self-regulatory and programmed/academic factors are of limited value.

Similarly, the results of the analysis of variance of differences between pooled FT and NFT means on the 41 classroom process variables are of limited value. The 13 variables for which significant differences were found are helpful in determining how FT differ from NFT classes, but they do not help identify significant treatments or treatment components.

What seems more useful in determining the degree to which particular sponsors differ from NFT are the results of the multiple range test which compares separate sponsors' means with the mean of NFT classrooms on each of the 41 classroom process variables.

F. Results:

1. Consistency of sponsor implementation. The following inter-project differences on each of the five implementation factors were reported. (See table on following page.)

   "Table 103 summarizes the results of judgmental interpretation of factor scores across classrooms within and between projects for each of the seven sponsor categories in which observation occurred in two different projects. This table shows that for the self-regulating and programmed academic factors, apparent differences between groups of classrooms in two different projects were small for nearly all sponsors. In contrast, most of the seven sponsors show higher variability across two groups of classrooms for the remaining three factors" (p. 299).

2. Discernible differences among sponsors.

   a. Sponsor classroom scores were pooled for each sponsor on the self-regulatory and programmed/academic factors. Sponsors were graphically contrasted. (See graph on page A-29.)

   b. SRI reports the following results of a multiple range test of sponsor means on 41 classroom process variables. These results indicate inter-sponsor differences as well as sponsor/NFT differences. We present SRI's table in its entirety. (See pages A-31 and A-32.)

INTER-PROJECT DIFFERENCES WITHIN SPONSOR CATEGORIES FOR FACTOR SCORES**

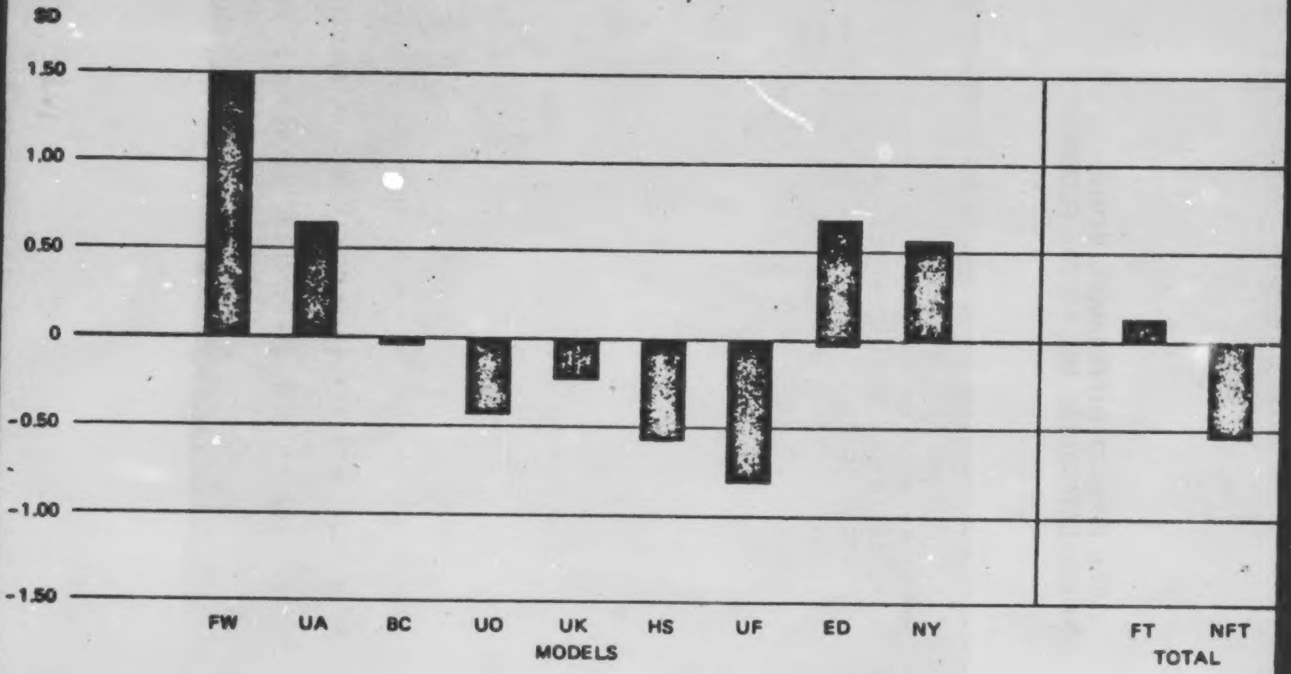
Sponsor	Factors				
	Self-Regulatory	Child-Initiated Interactions	Programmed/ Academic	Expressive	Child Self-Learning
FW	-	*	-	*	-
UA	-	-	-	*	*
BC	-	*	-	-	*
UO	-	*	-	*	*
UK	-	*	-	*	*
HS	-	*	*	*	-
UF	*	-	-	-	*

* = apparently large inter-project differences within sponsor categories.

- = apparently small inter-project differences within sponsor categories.

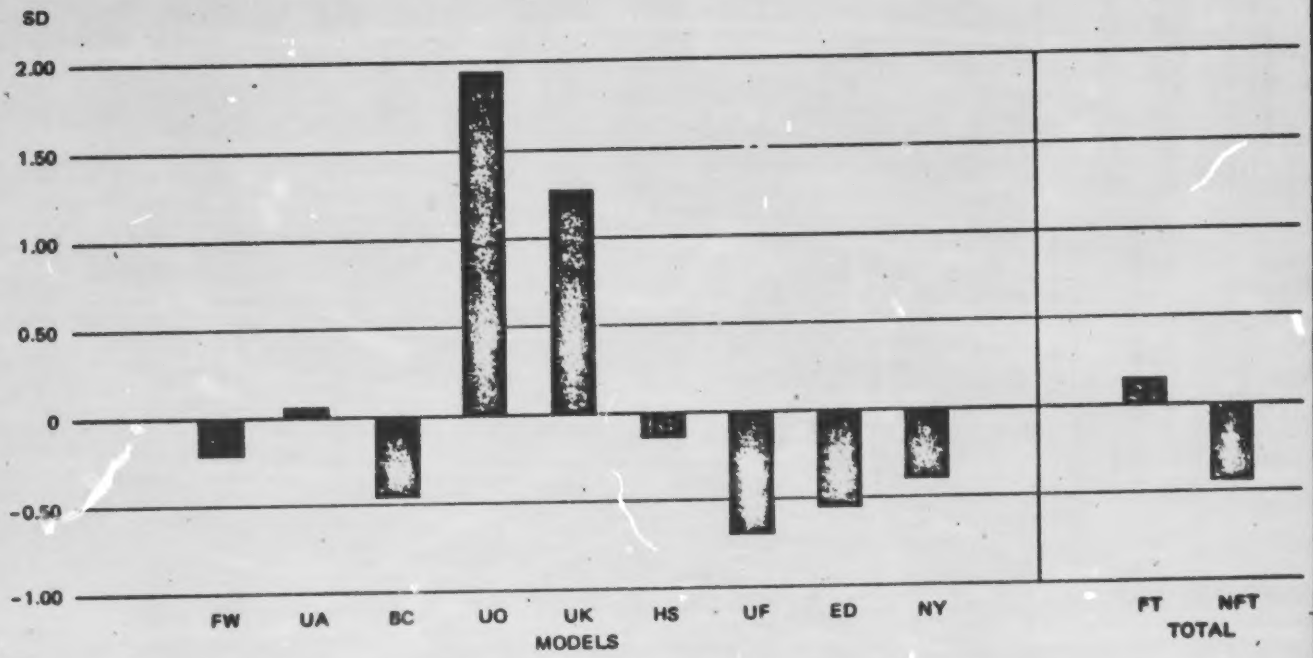
**Table 103, p. 300.

SPONSOR SCORES FOR THE SELF-REGULATORY FACTOR*



[redacted]gur[redacted] p.[redacted] .

SPONSOR SCORES FOR THE PROGRAMMED ACADEMIC FACTOR*



SRI observes: "It is difficult to avoid an impression of intersponsor differences from these graphs" (p. 300).

---

*Figure 10, p. 302.

TABLE 103

OBSERVED SPONSOR DIFFERENCES FOR CERTAIN CLASSROOM PROCESS VARIABLES

PROCESS VARIABLES	SPONSOR MEAN* LOW — HIGH
Activities (CCL)	
1. A-Lunch, snack	UO, UA, ED, UK, HS, UF, NY, BC, FW
2. B-Group time, story, singing, dancing	UO, UK, NY, UF, BC, HS, UA, FW, ED
3. C-Arithmetic, numbers, math., reading, alphabet, language development	ED, UF, BC, HS, FW, NY, UA, UK, UO
4. D-Social studies, geography, science, natural world	UO, UF, UK, NY, BC, UA, FW, HS, ED
5. E-Games, puzzles	UO, UF, UK, UA, HS, BC, NY, ED, FW
6. F-Arts, crafts, sewing, cooking, pounding, sawing	UF, UK, NY, UO, HS, BC, FW, UA, ED
7. G-Blocks, trucks, dolls, dress-up	UO, UF, NY, UK, HS, UA, FW, BC, ED
8. Adult with small groups in academic activities	ED, UF, FW, BC, NY, HS, UA, UK, UO
9. Academic activities	ED, UF, HS, BC, FW, NY, UA, UK, UO
10. Independent child activities	UF, UO, HS, UK, BC, UA, NY, ED, FW
11. Wide variety of activities	UF, UO, NY, BC, HS, UK, UA, FW, ED
12. Adult with one or two children in all activities	HS, NY, UO, UF, UA, UK, BC, FW, ED
13. Aide participating in academic activities	ED, NY, BC, HS, UF, FW, UA, UO, UK
Interactions (FMO)	
14. Adult informing child symbolically	NY, UA, UO, HS, FW, UF, BC, UK, ED
15. Adult asking child a direct question	FW, UF, BC, UK, ED, HS, NY, UA, UO
16. Direct question followed by child response	FW, UF, BC, UK, ED, HS, NY, UA, UO
17. Adult praise and corrective feedback	UF, ED, UA, FW, NY, BC, HS, UK, UO
18. Child response followed by adult feedback	ED, UK, UA FW, NY, BC, UF, HS, UO
19. Adult informing child	FW, NY, HS, ED, BC, UA, UF, UK, UO
20. Adult asking child thought-provoking questions	UO, NY, ED, FW, UK, BC, UF, UA, HS

↓ = Mean for aggregate of NFT classrooms observed (N = 26).

*Sponsor means on each variable are ordered in increasing magnitude. Underlining indicates subsets of no significant difference (p <.05) between sponsors, as determined by the multiple range test, Newman-Keuls method.

43

TABLE 105 (Concluded)

PROCESS VARIABLES	SPONSOR MEAN LOW HIGH
21. Adult informing child with concrete objects	UO, NY, ED, BC, UA, UF, FW, HS, UK
22. Adult acknowledgment	ED, UK, FW, UF, UA, UO, BC, NY, HS
23. Child informing himself with objects	UK, FW, BC, UF, UA, ED, HS, NY, UO
24. All child's self-learning	NY, UF, UA, FW, BC, ED, HS, UK, UO
25. Child informing another child	ED, UO, UF, NY, UK, BC, FW, HS, UA
26. Child informing himself symbolically	NY, UA, UF, FW, BC, ED, HS, UO, UK
27. Child questioning adult	UK, UF, UO, BC, ED, NY, HS, FW, UA
28. Child's self-expression	UO, UK, ED, HS, UF, BC, UA, FW, NY
29. Adult communication focus--one or two children	ED, NY, UF, FW, HS, UO, UA, BC, UK
30. Adult communication focus--small group	FW, BC, UA, NY, UF, HS, UK, ED, UO
31. Adult communication focus--large group	UO, HS, NY, UK, UA, FW, BC, ED, UF
32. Adult praise/acknowledgment of child	ED, UA, FW, UF, BC, NY, HS, UK, UO
33. Adult positive corrective feedback	UF, ED, HS, NY, UA, BC, UK, UO, FW
34. Adult negative corrective feedback	UF, UO, ED, UK, FW, NY, HS, UA, BC
35. Adult negative affect	UO, FW, NY, UA, ED, UK, HS, BC UF
36. Child negative affect	UO, ED, UF, BC, UA, NY, FW, UK, HS
37. All negative affects	UO, ED, UF, BC, UA, NY, UK, FW, HS
38. Adult to child positive affect	NY, BC, UO, ED, UK, UA, UF, HS, FW
39. Child to adult positive affect	UO, UF, NY, ED, BC, UK, UA, FW, HS
40. All positive affect	ED, NY, BC, UK, UO, UF, UA, HS, FW
41. Child positive affect	ED, UF, UO, BC, HS, NY, UK, UA, FW

3. Discernible differences between FT and NFT classroom activities.

   a. The graphed differences between pooled FT and NFT scores on the self-regulatory and programmed/academic factors are contained in the figures on pages A-29 and A-30. On the basis of visual inspection, there is an obvious difference between FT classrooms taken as a whole and NFT classrooms.

   b. Pooled FT and NFT scores on 41 classroom process variables were tested for significant ($p < .05$) differences. Significant differences were identified in 13 variables. In Table 104 below (p. A-34), SRI presents these variables in terms of the degree to which they are probably related to adult-to-child ratios in the classrooms in the following manner.

   c. Multiple range testing of the differences among sponsor and NFT means on classroom process variables produced the results of Table 105 (p. A-35). SRI notes that, "All sponsors are below the NFT mean on teacher didactic behaviors." We would also note that:

      (1) FT sponsors rank above NFT on providing a lunch or snack, adults working with small groups in academic activities, adult communication focus on small groups, and

      (2) All but one sponsor ranks above NFT on the number of games and puzzles, a wide variety of activities, an aide participating in academic activities, and child informing himself with objects.

      Additional overall FT/NFT as well as sponsor/NFT differences can be learned by examining the table.

Table 104

PROCESS VARIABLES THAT DIFFERENTIATE* BETWEEN FT AND NFT CLASSROOMS

Relation of Process to Adult-to-Child Ratio in the Classroom	Characteristic Processes	
	Follow Through	Non-Follow Through
Almost certainly directly related to number of adults	Adult with one or two children in academic activities Adult with small groups of children in academic activities Aides participating in academic activities Adult communication focus on small groups of children	Adult communication focus on large groups of children Adult informing children
Probably related to number of adults	Arts, crafts, sewing, cooking, pounding activities Blocks, trucks, dolls, dress-up activities Guessing games, table games, puzzles Independent child activities Snacks, lunch, group time Wide variety of activities	
May not be related to number of adults	Academic activities	

---

*Differences between FT classrooms (N = 97) and NFT classrooms (N = 26) were statistically significant ($p < .05$).

From p. 304 of the report.

III. PARENT-FOCUSED INQUIRIES

A. Research Questions: Questions were formulated at the same three levels of analysis as were used for child-focused inquiries.

1. National impact of the program. "On the average, is Follow Through as a national program of intervention producing measurable impact?" (p. 76).

2. Sponsor-level impact of the program. Have respective planned variations taken as a whole produced a measurable impact?

3. Site-level impact of the program. Have particular planned variations employed at each of the sites had a measurable impact?

B. Sample:

1. Method of construction: The list from which parents were drawn was composed of parents of children from the samples used in child-focused studies described above. From this list, households were selected on the basis of the following two criteria:

   -- "At sites where 110 or fewer entering pupils were tested, all corresponding households were scheduled for interviews.

   -- At sites where the number of pupils tested exceeded 110, random selections of 110 households defined the interview sample."

2. Description: Parents were interviewed in the spring of their children's first year in FT--i.e., Spring, 1970, for Cohort I and Spring, 1971, for Cohort II. The number of parents interviewed each year was 14,282 (96.5% of the sample selected) in 1970, and 14,778 (97.5% of the sample selected) in 1971 (p. 51).

C. Variables:

1. Independent: Being a FT or a NFT parent.

2. Dependent: Four parent impact measures were used (pp. 61-62):

   -- "Parent-child interactions, or the extent to which parents report they actively interact with their children in such activities as talking with their children, taking their children on trips, helping their children with school work, reading to them, accepting assistance from them, and acknowledging their progress in school.

-- "Parent-school involvement, or the extent to which parents report they are actively participating in various school-related activities, such as classroom visits, volunteer assistance, parent/school meetings, and external contacts with school personnel.

-- "Child academic expectations, or the extent to which the parent reports satisfaction with the child's progress and optimism regarding the child's future, both academic and non-academic. (E.g., what are the child's expected grades, chances of getting a good job, chances of going on to college?)

-- "Sense of control, or the extent to which the parent reports a sense of concern and control over school procedures, educational reforms, and school awareness of and responsiveness to parent and community desires and needs."

All four measures were constructed by selecting "judgmental grouping" of items on the parent interviews. "Responses to each such item were tabulated and distributed, and the resultant distributions were each dichotomized at their respective medians" (p. 61). Individual scores lying below the median of a given item were assigned a value of 0; scores above the median were assigned a value of 1. Then the rescaled items for each group were summed, and the summed scores were standardized.

Although site-level and sponsor-level findings are reported for each of these four variables, findings at the national level contain a fifth--"POS. EVAL. OF CHILD LEARNING." Curiously enough, this fifth parent measure is not defined anywhere in the report. Its meaning is thus left somewhat up in the air. In the absence of any explanation of this variable, we do not deal with it further.

3. Covariates: The following ten covariates were used (pp. 84-85):

"(1) No. classroom units--number of parent groups of data, aggregated in terms of the corresponding pupils' classrooms.

(2) Av. parents/classroom grp--average number of parents per unit meeting the completeness of data requirement.

(3) % Parents w/o HS diploma--the percentage of parents of pupils in the class who DO NOT have high school diplomas, averaged across classrooms.

(4) % Parents w skilled occup.--the average percentage of parents of pupils in the class who have occupations classified as 'skilled'--i.e., professional, clerical, manager, sales, craft, etc., but NOT service, laborer, operative, housewife, etc.

(5) % Pos. eval. of child lrng--satisfaction rating of parent ('very' to 'not') with child's progress in school; percentage of 'very satisfied' responses.

(6) % Parents Black--average percentage of pupil parents who are Black.

(7) % Reporting use of preschool--percentage stating their FT/NFT child participated in some preschool program.

(8) % Poverty eligible--average percentage meeting OEO poverty guidelines as shown in following tabulation.

(9) % Head of household employed--the average percentage of heads of households who report they are currently employed.

(10) % Head household male--the average percentage of heads of households who are male."

D. Analytical Method:

1. Overall strategy: As was the case with the child-focused analysis, unintended differences between comparison groups were known to exist. The analytical method used would have to be suitable for adjusting for these differences.

2. Model: "Analysis of covariance (ANCOVA) was the method chosen for integrating data from the various sources and performing tests of the major and sub-hypotheses. The specific ANCOVA model employed is a fixed effects, one-way design in which approaches (planned variables) are arranged as the treatment variables, and control variables are used as covariates. Furthermore, within each treatment group in the one-way design, the FT project samples and the corresponding NFT control groups are nested. This analysis design was implemented separately for pupil, parent, and teacher variables and for each cohort sample, grade stream, and assessment interval. Thus, 3 (pupil, parent, teacher) x 2 (Cohort I, II) x 2 (K, EF), or 12 independent analyses were performed on 1971 data, and 2 additional analyses were performed on the 1970 K and EF pupil data. Analyses of teacher and parent variables were not performed on the 1970 Cohort I subset" (p. 76).

E. Comments:

1. Variables: The report does not discuss the reliability or validity of either the dependent variables or the covariates. What was actually being measured or how consistently whatever was being measured was measured remains unclear.

   More specifically, the judgmental groupings of items on the parent interview are not explained in sufficient detail to permit an informal assessment of the validity of the scores based on these groupings.

More seriously, it is difficult to specify what the highly processed parent scores really mean. The process of constructing the parent scores involved the following steps. First, a set of concepts, presumably reflected by the labels attached to the respective scores, was chosen to group items on the parent questionnaire. Second, responses on particular items were rescaled to a (0,1) scale based on a median division of the distribution of each response. Third, the rescaled responses for the set of items within each group were summed. And, fourth, the summed rescaled scores were standardized.* Although the numerical values generated by this process still wear the labels that informed the initial grouping of items, it is difficult to establish any clear logical link between the numerical values and the concepts implied by the words of the labels.

2. Analytical Method: The method used for analyzing parent scores and the manner in which results are reported are the same as those used for the child-focused outcome inquiries. They are thus susceptible to the same reservations and criticisms laid out in detail in earlier comments. Interpretation of the parent-focused results is encumbered by an even more serious problem.

   SRI's clear intent was to identify the measurable impact or effects of the FT program upon parents. Yet, it attempts to draw causal inferences on the basis of data collected at only one point in time--during the spring of the child's first year of FT experience. Causal inferences based on post-hoc only assessment in such a non-experimental setting are obviously highly hazardous. The difference between FT and NFT parents is more appropriately interpreted as just that--a difference noted at a single point in time--and not as a product of the FT program. The results reported by SRI concerning parent "outcomes" can only be used to support the argument that FT has or has not had a measurable impact on parents if one adopts the untenable assumption that FT and NFT parents were equivalent at the start of the FT intervention.

F. Results: Results are reported with reference to the three levels of analysis incorporated in the research questions: national, sponsor, and site-level.

   1. Results at the national or overall level (pp. 275-293):

      a. Cohort I-EK: "The parent outcome measures are . . . not especially noteworthy, except for the parent/school involvement variable, for which 81 percent of the projects display FT-favoring results. This suggests that FT is having an impact on the degree to which parents become involved in school-related activities. (One must keep in mind that these measures are obtained during the child's first year of FT participation. Thus, parent outcomes are first-year outcomes, regardless of the cohort or grade stream.)" (p. 280).

---

*For analysis purposes a fifth step was undertaken, the calculation of group mean differences, of standardized, summed rescaled responses.

b. Cohort I-EF: "The parent outcome results for this cohort sample show a provocative and somewhat paradoxical pattern. Overall results showed FT-favoring differences (both average project effects and percentage of FT-favoring differences) for the parent-child interactions, the parent-school involvement, and the sense of control variables. The average project effect was NFT-favoring for parent expectations; in only 20 percent of the projects did FT parents report higher expectations for their child's success than did NFT parents. But, ironically, 80 percent of these same projects showed higher proportions of FT parents reporting they were satisfied with their child's current status. This might indicate that FT parents in these predominantly Southern rural projects simply have lower overall aspirations for their children and, hence, appear satisfied with their children's current progress. But this interpretation is hard to reconcile with the more positive FT results for the school involvement and sense of control measures, unless these parents are responding to the expectation measures on the basis of a larger socio-cultural context. (Many of these projects involved very poor Black FT families and less poor non-Black families.)" (p. 282).

c. Cohort II-EK: "Encouraging effects are also present for the parent outcome variables, where four of the five measures show overall FT-favoring differences (the parent expectation variable is the exception). For all variables, the majority of the projects showed FT-favoring results" (p. 282).

d. Cohort II-EF: "Since only four projects are included in this sample, summary statistics are likely to be unreliable" (p. 283). FT/NFT differences at these four sites did, however, favor FT on the four parent outcome measures.

2. Results at the sponsor level: Results are reported for CI-EK, CI-EF, DII-EK, and CII-EF parents. However, no sponsor had a sufficient number of sites within each cohort-grade stream to warrant analyses for all four types of parents. Effects are reports in terms of adjusted FT/NFT differences on each of the four parent outcome measures. Statistical tests were performed to determine whether these differences were significantly different from 0.0 at the .05 level of confidence.

a. Far West Laboratory for Educational Research and Development (pp. 108-111)

  (1) Cohort I-EK. A significant difference was obtained on one of the four outcome measures--parent-school involvement.

  (2) Cohort II-EK. A significant difference was obtained on a single outcome measure--parent-child interventions.

b. University of Arizona: Tucson Early Education Model (pp. 130-132)

  (1) Cohort I-EK. Significant differences were detected on two of the outcome measures: parent-school involvement and child academic expectations.

  (2) Cohort I-EF. Significant differences were obtained on parent-school involvement and sense of control.

c. Bank Street College of Education Approach (pp. 148-149)

Cohort I-EK. A significant difference was obtained on parent-school involvement.

d. University of Georgia, Mathemagenic Activities Program (pp. 153-155)

Since there was only one site from which data were collected for this analysis, the ". . . risk of faulty interpretation is considered very high." Analysis based on this single site was for Cohort I-EF parents. No significant differences were found on any of the parent outcomes.

e. University of Oregon, Engelmann/Becker Model for Direct Instruction (pp. 171-175)

  (1) Cohort I-EK. Significant difference on parent-school involvement.

  (2) Cohort I-EF. No significant differences were obtained.

f. University of Kansas, Behavioral Analysis Approach (pp. 186-187)

Cohort I-EK. The only significant difference was on parent-child interactions. In this case, the difference was NFT-favoring. NFT parents and children appear to interact more. However, SRI notes that, "Because of missing data and the lack of comparability of the comparison groups, these results are probably invalid."

g. High/Scope Educational Research Foundation, Cognitively Oriented Curriculum Model (pp. 198-200)

Results were based on analysis of a single site since serious data problems were encountered elsewhere. This was Cohort I-EF site. Three significant FT-favoring differences were found: parent-child interactions, parent-school involvement, and parents' expectations concerning their children's academic futures.

h. University of Florida, Florida Parent Education Model (pp. 215-217)

Cohort I-EK. No significant differences were observed.

i. Education Development Center, EDC Open Education Program (pp. 233-235)

Cohort I-EK. A single, significant difference indicated that NFT parents and children interact more than FT parents and children do.

j. New York University, Interdependent Learning Model (pp. 243-246)

Cohort I-EK. Two, significant FT-favoring differences were obtained: parent-school involvement and parents' sense of control.

k. Southwest Educational Development Laboratory (pp. 250-251)

Cohort I-EK. No significant differences were obtained. SRI notes, however, that since all the data are taken "from a single sample within a single project, we feel no conclusions can be justified at this time."

3. Results at the site level: As was the case with the child outcome results reported earlier, a dilemma is confronted in attempting to summarize the parent "outcome" findings--i.e., to extract outcome findings from the contextual information needed to sensibly interpret them or to report the findings and all the situational detail. We have resorted to the same, less than satisfactory procedure of reporting site-level differences on the four parent measures along with SRI's estimation of how well FT and NFT groups were matched.*

The following four tables report adjusted FT/NFT differences according to the respective cohort-grade stream groups.

---

*See Child-Focused Inquiries for an explanation of how these estimations were formed.

(We have deleted reported differences on a fifth parent measure--"POS. EVAL OF CHILD LEARNING"--on the grounds that it has not been defined.)

"The left most column contains 'project codes.' These stand for:

SS--Self sponsored project	HS--High/Scope
FW--Far West Laboratories	UF--University of Florida
UA--University of Arizona	ED--Education Development Center
BC--Bank Street College	NY--New York University
UO--University of Oregon	SW--Southwest Educational Development Laboratory
UK--University of Kansas	
PI--Parent implemented	

(a), (b), (c), (d) refer to particular sites of a given sponsor."

SUMMARY OF SITE-LEVEL DIFFERENCES ON FOUR PARENT MEASURES FOR COHORT I-EK*

		PARENT MEASURES			
PROJECT CODE	MATCH	PARENT-CHILD INTERACT.	PARENT-SCHOOL INVOLVEMENT	CHILD ACADEMIC EXPECTATION	SENSE OF CONTROL
TWO YEAR OUTCOMES					
SS(b)†	GOOD	.373	.356	.286	.644
SS(c)†	MOD	-.038	-.188	.019	-.368
SS(d)†	GOOD	.181	.002	.162	.438
SS(e)†	POOR	-.070	-.426	-.337	-.515
FW(a)	GOOD	.240	.091	-.535	-1.12*
FW(b)†	GOOD	-.147	.721*	.112	-.047
FW(c)	MOD	-.287	-.091	-.800	.115
UA(a)	GOOD	.146	.368	.197	.015
UA(d)†	MOD	.460	.697*	.589	-.437
BC(a)	POOR	—	—	—	—
BC(b)	MOD	-.157	.665	.185	.288
BC(c)	MOD	.408	.735	-.367	.438
BC(e)	MOD	.595*	.328	-.085	-.285
UO(a)	POOR	-.192	.303	-.560	.160
UO(b)	MOD	.503	.304	-.025	-.022
UO(c)	GOOD	-.031	.156	.423	-.252
UK(a)	MOD	.041	.115	.210	-.264
UK(b)†	POOR	-.157	.653	-.446	.217
UK(c)†	GOOD	.492	.023	-.132	.372
HS(b)	MOD	-.029	.058	-.405	-.135
UF(a)	MOD	-.119	.084	.144	.146
UF(c)†	GOOD	.437	.505	-.060	-.069
ED(b)†	GOOD	-.216	-.225	.192	.203
ED(c)†	GOOD	—	—	—	—
NY(a)	GOOD	-.011	.174	-.395	.341
NY(b)	GOOD	-.144	.291	-.333	.348
SW	MOD	.080	-.324	-.079	.181
PI†	MOD	.076	.310	.534	-.815

---

*Taken from Table 96 of SRI's report, p. 276.

SUMMARY OF SITE-LEVEL DIFFERENCES ON FOUR PARENT MEASURES FOR COHORT I-EF*

		PARENT MEASURES			
PROJECT CODE	MATCH	PARENT-CHILD INTERACT.	PARENT-SCHOOL INVOLVEMENT	CHILD ACADEMIC EXPECTATION	SENSE OF CONTROL
TWO-YEAR OUTCOMES					
SS(a)†	MOD	.723*	.832*	.260	.189
UA(b)	MOD	.279	.516	-.447	-.206
UA(c)†	MOD	-.394	-.066	-.028	.872*
BC(d)†	GOOD	.176	-.080	-.954*	.124
UG	MOD	.684	.065	-.578	.209
UO(d)†	POOR	-.038	-.312	-.349	-.102
UO(e)	--	--	--	--	--
HS(a)†	MOD	.776*	1.401*	.813*	.130
HS(b)	MOD	.281	-.349	-.364	-.791*
UF(b)	GOOD	-.241	.760*	-.115	-.341
ED(a)†	MOD	.237	.607*	-.227	.306

*p < .05

SUMMARY OF SITE-LEVEL DIFFERENCES ON FOUR PARENT MEASURES FOR COHORT II-EK**

		PARENT MEASURES			
PROJECT CODE	MATCH	PARENT-CHILD INTERACT.	PARENT-SCHOOL INVOLVEMENT	CHILD ACADEMIC EXPECTATION	SENSE OF CONTROL
FW(a)	GOOD	-.182	.643	-.034	-.195
FW(b)	GOOD	.797*	.718	.119	.411
FW(c)	MOD	-.050	-.623	-.542	.227
BC(c)	GOOD	.150	.036	.120	.273
UO(a)	MOD	1.036*	.595	.353	-.530
UF(a)	GOOD	.067	.352	.193	.379
ED(b)	GOOD	--	--	--	--
NY(a)	MOD	.309	.268	-.230	.357

*p < .05

---

*Taken from Table 98 of SRI's report, p. 278.

**Taken from Table 97 of SRI's report. p. 277.

SUMMARY OF SITE-LEVEL DIFFERENCES ON FOUR PARENT MEASURES FOR COHORT II-EF *

PROJECT CODE	MATCH	PARENT MEASURES: PARENT-CHILD INTERACT.	PARENT-SCHOOL INVOLVEMENT	CHILD ACADEMIC EXPECTATION	SENSE OF CONTROL
UA(c)	GOOD	.087	.101	.144	-.702
BD(d)	GOOD	.306	.456	.056	.179
UO(d)	MOD	-.196	.277	-.100	1.57
UF(b)	POOR	--	--	--	--

---

*Taken from Table 99 of SRI's report, p. 279.

IV. TEACHER-FOCUSED INQUIRIES

A. Research Questions: Three research questions were addressed:

1. National impact of the program: Across the board, has FT had a measurable impact on teachers of FT classes?

2. Sponsor level impact of the program: Have respective sponsors' FT programs had a measurable impact on teachers?

3. Site-level impact of the program: Have particular planned variations, as implemented at each site, had a measurable influence on FT teachers?

B. Sample:

1. Method of construction: "During the last two weeks of April, 1971, questionnaires were mailed to all Follow Through coordinators, to be distributed to both Follow Through and non-Follow Through teachers whose classes had been tested in the Fall, 1970. During the last week of May, an additional 200 questionnaires were sent out in the Spring but not in the Fall. Each Follow Through teacher received one aide questionnaire for a regular classroom aide to fill out. If more than one aide worked in the classroom, the teachers were instructed to alphabetize the aides by the last name and give the questionnaire to the second one on the list" (p. 52).

2. Description: A total of 1774 (993 FT and 781 NFT) teacher questionnaires and 993 (presumably all FT) aide questionnaires were distributed in this manner. Of these, 1462 teachers and 804 aide questionnaires were returned.*

C. Variables:

1. Independent: Association with an FT or an NFT classroom.

2. Dependent: Only two outcome measures were used (p. 63):

"(1) Parent-educator image, or the extent to which teachers reported they felt it essential to 'get together with parents outside of the classroom' for purposes of

- Improving children's learning
- Improving classroom teaching
- Learning parents' views on teaching
- Improving school services to parents
- Improving school services to children
- Improving school services to community
- Parental understanding of school program.

---

*Comparative FT/NFT response rates are not reported.

(2) Professional acceptance of method, or the extent to which the teacher reports she would not prefer to adopt some teaching approach other than the one she is currently using."

The manner of which teacher variables are consistently treated as dependent variables is a bit of a puzzle in the report. While site-level and sponsor-level results are reported for the two dependent variables just mentioned, results at the national level also include FT/NFT contrasts on two additional variables--"job satisfaction" and "number of helpers." Summary commentary (e.g., p. 280ff) clearly refers to these two additional variables as dependent variables despite the fact that they serve as covariates for the site, sponsor, and national level analyses. For the remainder of the present discussion, we consider "parent-educator image" and "professional acceptance of method" as the only two dependent variables for the teacher analysis.

3. Covariates (p. 86):

"(1) No. of classrooms--number of classrooms for which data on teacher variables are sufficiently complete.

(2) Job satisf. rating--the average scaled satisfaction response (3 = very satisfied, 2 = satisfied, 1 = dissatisfied) regarding working conditions, i.e., equipment, supplies, classroom space, schedule, salary, and planning time.

(3) Book Resource scale--the number of book and library facilities available to the students (classroom, take home, central library, etc.).

(4) Race (Black/non-Black)--Black = 1; non-Black = 0.

(5) Ident. w community--the extent to which the teacher is both a long time member of the community and a resident of a neighborhood similar to that of the pupils.

(6) No. of helpers--the minimum number of helpers (aides or volunteers) untilized in the class.

(7) Able to choose assignment--the extent to which the teacher was able to choose her current school and classroom teaching assignment--(3 = teacher's own choice; 2 = at request of other; 1 = assigned; no choice).

(8) Trng & teacher experience--an aggregate training and experience variable indicating the extent of formal education, degrees held, specific course completions, certification, years of experience, grade levels taught, and tenure of the teacher (max. score = 12 points).

D. Analytical Method: The overall strategy and particular statistical model used were the same as for the child- and parent-focused analyses--i.e., a fixed effects, one-way ANCOVA.

E. Comments:

1. Sample: The information provided is less than adequate to determine how many of each type (cohort-grade stream) of teachers were included. The number of teacher questionnaires sent out and returned (993) clearly indicates that the teacher analytical subset is based on a larger collection of classrooms than that used for the child-focused analysis. That analysis was based on a sample of 582 classrooms.

   The difference between the two analytical subsets may be more than just the greater number of classrooms involved in the teacher analysis. It appears that the teacher analytic sub-sample also excludes some of the classrooms used in the child-focused analysis. Seven of the 28 Cohort I-EK sites, three of the eleven Cohort I-EF sites, and four of the eight Cohort II-EK sites do not have teacher results.

   The report acknowledges that "at present, no data are available on determinants of refusal" (p. 52). Moreover, no information is provided concerning the response rate of NFT teachers. In sum, it appears that the sample is an opportunity sample, the representativeness of which remains undetermined.

2. Variables: The appropriateness of the covariates used as well as the validity and reliability of all dependent variables and covariates remain undiscussed. The two dependent variables are constructed just as the parent outcome variables were. They are, thus, highly processed variables.

3. Analytical Method: The method used for analyzing teacher scores and the manner in which results are reported are the same as those used for the child-focused outcome inquiries. They are thus susceptible to the same criticisms and reservations laid out in detail in earlier comments.

   The analysis of teacher "outcomes" uses data collected at a single point in time. It is consequently impossible to legitimately attribute any observed difference to the program. Such differences are appropriately interpreted as merely the differences between two types of teachers.

   It is unclear if or how aide questionnaires were analyzed.

F. Results: Results are reported with reference to three levels of analysis: national, sponsor, and site-level.

1. Results at the national level (pp. 275-293):

a. Cohort I-EK: Conflicting results are obtained. While 76% of the sites report FT-favoring differences on professional acceptance of method, 86% report NFT-favoring differences on the other dependent variable, parent-education image.

b. Cohort I-EF: All differences favor FT on professional acceptance of method. Differences on parent-education image are evenly split. Only two differences are statistically significant and none is substantially significant.

c. Cohort II-EK: Differences at three of four sites favor FT on professional acceptance of the method. Differences on parent-educator image were divided two and two.

d. Cohort II-EF: Results for this cohort-grade stream were exactly the same as those for Cohort II-EK: Three of four sites reported FT-favoring differences on professional acceptance of method and two of four reported FT-favoring differences on parent-educator image.

The national level findings are consistent in the sense that,at a majority of sites of all cohort-grade streams, FT teachers appear to accept their planned variation method more than NFT teachers do.

2. Results at the sponsor level: Sponsor-level findings are summarized in the following table;"ss" indicates a statistically significant FT/NFT difference; "nss" indicates the difference is not statistically significant.

	Variables	
Sponsor	Parent-Educator Image	Professional Acceptance of Method
FW, I-EK	nss	nss
II-EK	nss	nss
UA, I-EK	nss	nss
II-EK	nss	ss
BC, II-EF	nss	nss
UG, I-EF	nss	nss
UO, I-EK	nss	ss
I-EF	nss	ss
UK	no results reported	
HS	no results reported	
UF, I-EK	nss	nss
ED, I-EK	nss	nss
NY, I-EK	nss	ss
SW, I-EK	nss	nss

3. Results at the site level: The following four tables report adjusted FT/NFT differences on the two teacher scores "parent-educator image" and "professional acceptance of the FT method" for 37 sites. These summary tables suffer from the same inadequacies noted earlier concerning child outcomes and parent measures. The many ways in which sites differ from one another have been reduced to a single reference, "MATCH"--how similar FT and NFT classes were on classroom composition and pupil family measurements. Consequently, ". . . it would not be unusual if teacher outcomes were unrelated to the quality of pupil/parent matches" (p. 284). In sum, the findings concerning these two teacher measures are best considered only as additional descriptive information concerning the differences among sites and not as program effects.

SUMMARY OF SITE-LEVEL DIFFERENCES
ON TWO TEACHER MEASURES FOR COHORT I-EK*

PROJECT CODE	MATCH	PARENT-EDUCATOR IMAGE	PROF. ACCEPT. FT
SS(b)†	GOOD	-.08	-.26
SS(c)†	MOD	—	--
SS(d)†	GOOD	-.05	.80*
SS(e)†	POOR	—	—
FW(a)	GOOD	--	--
FW(b)†	GOOD	.29	.40
FW(c)	MOD	-.23	.38
UA(a)	GOOD	-.05	.63*
UA(d)†	MOD	-.26	.23
BC(a)	POOR	-.88*	.20
BC(b)	MOD	-.41*	.29
BC(c)	MOD	-.04	-.22
BC(e)	MOD	-.38*	.51*
UO(a)	POOR	-.23	.88*
UO(b)	MOD	-.03	1.01*
UO(c)	GOOD	-.31	-.15
UK(a)	MOD	—	—
UK(b)†	POOR	—	—
UK(c)†	GOOD	—	—
HS(b)	MOD	—	--
UF(a)	MOD	-.34	.78*
UF(c)†	GOOD	-.79*	-.36
ED(b)†	GOOD	-.30	.22
ED(c)†	GOOD	.53*	.60
NY(a)	GOOD	-.16	.60*
NY(b)	GOOD	.03	.87*
SW	MOD	-.05	.21
PI†	MOD	-.61*	-.84

---

*Taken from SRI's report, Table 96, p. 276.

SUMMARY OF SITE-LEVEL DIFFERENCES ON TWO TEACHER MEASURES FOR COHORT I-EF*

PROJECT CODE	MATCH	PARENT-EDUCATOR IMAGE	PROF. ACCEPT. FT
SS(a)†	MOD	.49	.18
UA(b)	MOD	-.43	.72*
UA(c)†	MOD	.22	.37
BC(d)†	GOOD	--	--
UG	MOD	.10	.50
UO(d)†	POOR	.07	1.63*
UO(e)	--	--	--
MS(a)†	MOD	--	--
MS(b)	MOD	-.30	.12
UF(b)	GOOD	-.55	.62
ED(a)†	MOD	-.03	.04

SUMMARY OF SITE-LEVEL DIFFERENCES ON TWO TEACHER MEASURES FOR COHORT II-EK**

Project Code	Match	Parent-Educator Image	Professional Acceptance FT
FW(a)	GOOD	--	--
FW(b)	GOOD	- .36	1.08
FW(c)	MOD	.21	1.09
BC(c)	GOOD	--	--
UO(a)	MOD	.17	1.41
UF(a)	GOOD	- .38	- .11
ED(b)	GOOD	--	--
NY(a)	MOD	--	--

---

*Taken from SRI's report, Table 98, p. 278.

**Taken from Table 97, p. 277.

SUMMARY OF SITE-LEVEL DIFFERENCES ON TWO TEACHER MEASURES FOR COHORT II-EF*

Project Code	Match	Parent-Educator Image	Professional Acceptance FT
UA(c)	GOOD	.22	.46
BC(d)	GOOD	- .49	.93
UO(d)	MOD	- .46	- .03
UF(b)	POOR	.29	.11

---

*Taken from Table 99, p. 279.

1.2 Marvin G. Cline, N. Ames, R. Anderson, R. Bales, T. Ferb, M. Joshi, M. Kane, J. Larson, D. Park, E. Proper, L. Stebbins, and C. Stern. Final Report. Education as Experimentation: Evaluation of the Follow Through Planned Variation Model. Volume IA: Early Effects of Follow Through and Volume IB: Monographs. Cambridge, MA: Abt Associates, March 1, 1974.

Abt's 1974 report contains a number of separate analyses. The questions addressed by these analyses can be grouped into four types: child-focused, parent-focused, teacher-focused, and other inquiries. We have grouped our summaries of the analyses in the report according to these types:

CHILD-FOCUSED INQUIRIES

I. ONE-YEAR KINDERGARTEN OUTCOMES . . . . p. 57
II. SPONSOR VIGNETTES: WITHIN SPONSOR, ONE-YEAR EFFECTS . . . . p. 67
III. PRESCHOOL EFFECTS . . . . p. 76
IV. CHILD SEX DIFFERENCES . . . . p. 86
V. CHILD ETHNICITY . . . . p. 88
VI. ENTERING CLASS-ACHIEVEMENT LEVEL . . . . p. 90
VII. CLASS ETHNIC COMPOSITION . . . . p. 94
VIII. COHORT I THREE-YEAR EFFECTS . . . . p. 98
IX. INTER-COHORT ANALYSIS . . . . p. 100

PARENT-FOCUSED INQUIRIES

X. FT AND NFT, COHORT III PARENT CHARACTERISTICS . . . . p. 103
XI. FT AND NFT, COHORT I PARENT CHARACTERISTICS . . . . p. 106
XII. PARENTS' ATTITUDES AND BEHAVIORS . . . . p. 109

TEACHER-FOCUSED INQUIRIES

XIII. TEACHER CHARACTERISTICS . . . . p. 111
XIV. TEACHER TRAINING OF FT TEACHERS . . . . p. 114
XV. TEACHER ATTITUDES AND BEHAVIORS . . . . p. 117
XVI. TEACHER SATISFACTION WITH AND FAITHFULNESS TO SPONSOR'S MODEL . . . . p. 121
XVII. INTERRELATIONS OF TEACHER CHARACTERISTICS . . . . p. 124

OTHER INQUIRIES

XVIII. IMPLEMENTATION . . . . p. 128
XIX. A POTPOURRI OF ODDS AND ENDS . . . . p. 133

I. ONE YEAR KINDERGARTEN OUTCOMES (Vol. IA, pp. VII-1 to V-26; VII-31 to VII-44)

A. Research Questions:

1. "'Is there a global FT effect?' That is to say, when we merge all of the various sponsors into a single group and treat FT as an undifferentiated program, is there a detectable 'Follow Through' effect [after one year of kindergarten experience]?" (p. VII-2).

2. Have the respective sponsor models, considered separately, produced a detectable effect after one year of kindergarten experience?

B. Sample:

Data from two samples were used to address the research questions: a sample including big city schools (N = 288) and the same sample except with big city schools excluded (N = 255). Little discussion of how either were constructed is to be found in the report. "The criterion for inclusion in the studies at the various levels of analysis is the availability of sufficient numbers of valid scores on children, their parents, and their teachers" (p. III-2).

The larger of the two samples includes FT data from 37 big city schools.* The two samples were distributed across sponsors as follows (see pages A-55 and A-56 for distribution).

C. Variables:

1. Independent: one year of FT or NFT experience.

2. Dependent (pp. III-14 to III-16)

   a. "Wide Range Achievement Test (WRAT). This test is an individually administered measure of letter and number recognition, word reading, spelling, and oral and written arithmetic problems. Whereas national norms are available, spelling, reading, and arithmetic subtests were abbreviated for administration in the Follow Through evaluation. Thus, the published norms are not applicable to this modified version. The scores analyzed here are total raw scores with a range from 0 up to a maximum score of 84."

   b. "Peabody Picture Vocabulary Test (PPVT). This is an individually administered measure of picture recognition

---

*Abt compared results obtained in these two samples to determine FT effects in big city schools. Such a comparison is a roundabout and weak method of investigating this issue. Consequently, we report the results from the sample including big city schools here as though they were the products of a replication of the analyses performed on data from the smaller sample.

1. The smaller sample which excludes big city schools.*

Distribution of the FT and NFT Schools Across Ten Sponsors For the Study of One Year Kindergarten Effects

	2	3	5	7	8	9	10	11	12	14	TOTALS
FT	29	21	12	11	16	8	15	10	9	6	137
NFT	20	20	10	9	12	9	9	9	11	5	114
TOTALS	49	41	22	20	28	17	24	19	20	11	251

---

*Throughout the Abt report and these summaries the following code numbers are used for the various sponsors:

2 - Far West Laboratory
3 - U. of Arizona
5 - Bank Street College
7 - U. of Oregon
8 - U. of Kansas
9 - High/Scope
10 - U. of Florida
11 - Educational Development Center
12 - U. of Pittsburgh
14 - Southwest Developmental Lab.

2. The larger sample including big city schools.*

Distribution of the Complete Analytic Subset
of FT and NFT Schools Across Ten Sponsors

	2	3	5	7	8	9	10	11	12	14	TOTALS
FT	29	21	16	11	20	11	17	13	9	9	156
NFT	20	20	15	10	15	12	9	12	11	8	132
TOTALS	49	41	31	21	35	23	26	25	20	17	288

This sample represents the following changes from the previously described sample.

	2	3	5	7	8	9	10	11	12	14	TOTALS
FT	0	0	+4	0	+4	+3	+2	+3	0	3	19
NFT	0	0	+5	+1	+3	+3	0	+3	0	3	18
											37

---

*Table VII-6, p. VII-32.

vocabulary. The child is required to point to the correct picture corresponding to the word spoken by the examiner. The test was administered according to standard procedures; however, a number of pictures were modified to reflect Black rather than White ethnic characteristics. Thus, the test scores are not directly parallel to those derived from the standard version. Total scores equal the number of correct responses.

c. "Metropolitan Achievement Tests (MAT). Three subtests from this multiple choice, group-administered test were analyzed as dependent variables: Reading, Listening for Sounds, and Arithmetic. The first is a test of letter and word reading. The second is a test of recognition of initial, medial, and final word sounds, representing important reading readiness skills. The Arithmetic subtest is a measure of basic math concepts, vocabulary, and addition and subtraction skills."

d. "Gumpgookies. This test, developed by Adkins and Ballif, measures the child's motivation to achieve in school. The hypothetical constructs underlying this measure are theorized to be a dynamic interaction of learned responses unrelated to intellectual ability, including: the child's knowing and performing activities directed toward achievement; enjoyment of the school situation; self-evaluation; self-confidence in physical activities; and the child's purposive behavior toward accomplishing future goals. As an individually administered picture test, the child is required to point to one of two semi-projective 'Gumpgookie' figures which presumably reflects the orientation of the child either toward or away from one of the five dimensions outlined above. The Gumpgookies are vague figures with the outline of a head, arms, and legs, resembling 'Casper' the ghost.

Total raw test scores were analyzed for these studies, the maximum possible being 60."

e. "Locus of Control (Locus-positive and Locus-negative). This individually administered picture test, developed by Shipman, reflects the child's perception of the extent to which he, or others in his environment, are responsible for the events that happen to him--a hypothetical dimension first explored by Rotter and later substantiated by a wide body of research. The events characterized in this test are restricted to the academic and social situations in the child's school life. Two scores are derived from the test Locus-positive

(11 items), reflecting the child's perceived responsibility for good events; and Locus-negative (9 items), reflecting the child's responsibility for unfavorable happenings in his school situation. Research on this test suggests that different intra-personal dynamics underlie these two scores; thus the total Locus of Control score is not separately analyzed here.

* * * * * *

f. "The number of days a child missed school throughout the kindergarten year was analyzed as an indirect measure of a general capacity for a Sponsor to attract pupils consistently. While children who enjoy school are more apt to attend, this interpretation cannot be fully accepted without qualifications concerning other reasonable determinants of absence such as sickness, parental factors, or environmental phenomena operating within the community."

3. Covariates: Outcome contrasts were adjusted using the following eleven covariates (p. VII-8):

   a. "Wide Range Achievement Tests (WRAT), administered as an achievement pretest in Fall, 1971, and expressed as a school mean of child scores in participating FT or NFT classes.

   b. Percentage of Black pupils in the school's study populations. Percentage of minority pupils, including Indians, Orientals, and Spanish-speaking children as well as Blacks.

   c. Years at current address, according to information provided in parent interviews, averaged to school level.

   d. Adjusted income level, the school average of a composite prosperity index which incorporates parent-reported income, family size, and whether or not the family is located in a rural area.

   e. Mothers' education, for each school, the percentage of mothers reporting at least a high school education.

   f. Parent-school receptivity, a composite index, discussed in detail in Monograph I, reflecting the extent to which parents perceive that their child's school welcomes parental participation.

   g. Western region, one (1) if the school is located in the West; zero (0) otherwise.

h. Southern region, one (1) if the school is located in the South; zero (0) otherwise.

i. Metropolitan, one (1) if the school is located within a Standard Metropolitan Statistical Area (SMSA); zero (0) otherwise.

j. Middle-sized cities, one (1) if the school is located in a metropolitan community whose population falls between 50,000 and 200,000; zero (0) otherwise."

D. Analytical Method: ANOVA and ANCOVA models were used to identify FT/NFT contrasts at both the overall level and at the sponsor level. Nested designs were used to calculate sponsor effects. Factorial designs were used to calculate overall FT effects. The school was used as the unit of analysis for results reported here. Unadjusted and adjusted FT/NFT differences were expressed in terms of standard deviations of each of the eight dependent variables. Abt's report recommends .25 standard deviations as an indicator of "a conceptually interesting level of effect" (p. V-8).

E. Comments: Chapter 3 of the main text of this summary report discusses in detail a number of common methodological problems confronting the overall FT evaluation effort. The analyses of kindergarten effects are vulnerable to many of these problems. Consequently, the results reported below need to be read with a healthy measure of skepticism and caution.

In addition to general methodological problems, a handful of specific items requires special mention.

1. Sample: The opportunity samples drawn provide no assurance of equivalent testing situations across sponsors, equivalent testing situations for FT or NFT groups, nor the representativeness of analytic samples with respect to some larger population. Consequently, as the authors of the report note, ". . . our analytic subsets [we hesitate to call them samples] are not representative of any definably larger population" (p. II-7).

2. Variables: The report presents the following Hoyt internal consistency coefficients for seven of the eight dependent variables (p. III-14):

WRAT		.92	Gumpgookies		.88
MAT:	Reading	.83	Locus of control:	positive	.43
	Listening	.86		negative	.22
	Arithmetic	.87			

Most are acceptable by conventional standards. However, the low reliability of both locus of control measures undermines the confidence one can place in findings concerning them.

The fallibility of the covariates is not examined. In the absence of evidence concerning the fallibility of these measures, it is difficult to say what adjustments have actually been made. In turn, it becomes difficult to interpret "adjusted" FT/NFT differences or lack thereof.

3. Analytical Method: Abt's report explicitly states that no weighting procedures were employed to account for differences in school enrollment. Similarly, differences in class sizes were not taken into account (p. III-2). Both school and class sizes did vary:

Schools		Classes	
5 to 10 children	8%	5 to 10 (approx.)	45%
11 to 15 (approx.)	17%	11 to 14	35%
16 to 124	75%	15 to 24	20%

In both cases, then, the analysis attributes as much importance to a unit of analysis with few children as it does to a unit with many more children.

F. Results: Results of the analysis based on the smaller sample are reported first, followed by those of the analysis based on the larger sample.

Based on Sample of Big City Schools

1. Overall FT/NFT contrasts indicate that:
   - a. Both unadjusted and adjusted differences are in the desired direction on all eight outcome measures.
   - b. Adjusted differences are all slightly greater than the unadjusted.
   - c. The magnitude of adjusted differences ranges from a low of 0.117 standard deviations for locus of control/position to 0.413 standard deviations for Gumpgookies.
2. FT/NFT contrasts at the sponsor level indicate that, "At the kindergarten level, Sponsors show every conceivable pattern across achievement, motivations and absence measures, and no two Sponsors show the same pattern" (p. VII-30). The particular contrast differences are presented in the following table (p. A-61).

Adjusted FT/NFT Differences Based on Pooled School Means for Respective Sponsors*

Sponsors (no.FT/no.NFT schools)	OUTCOME MEASURES							
		MAT				Locus of Control		
	(1) WRAT	(2) Listening	(3) Reading	(4) Numbers	(5) Gumpgookies	(6) Posit.	(7) Neg.	(8) Absence
2. Far West (29/20)	.023	-.182	.283	-.064	.607	-.160	-.047	.083
3. U. of Arizona (21/21)	.274	-.223	.159	-.463	.501	-.263	.160	-.150
5. Bank St. College (12/10)	-.259	-.685	-.493	-.495	.631	.392	-.120	-.114
7. U. of Oregon (11/9)	.756	1.425	.604	.999	-.015	-.055	.448	-.247
8. U. of Kansas (16/12)	1.289	1.006	.858	1.387	.435	-.276	-.006	-.470
9. High Scope (8/9)	.585	.453	.842	.634	.664	.387	.092	-.781
10. U. of Florida (15/9)	.316	.194	.406	.091	1.08	.233	.771	-.238
11. EDC (10/9)	-.031	-.245	-.199	-.117	.039	.040	-.099	-.570
12. U. of Pittsburgh (9/11)	.645	.0008	.113	.969	1.152	.314	1.033	-.159
14. SEDL (6/5)	.510	-.124	.430	.103	-.965	.557	.152	-1.050

---

*Differences are expressed in standard deviations of each variable. Positive values indicate that FT adjusted means were higher than NFT scores. Positive differences are interpreted as desirable for all but the last outcome measure, absence, for which a negative difference is preferred. (Taken from Figures VII-2 to VII-11.)

Based on Sample Including Big City Schools

1. Overall FT/NFT contrasts:

   -- confirm the results obtained with data from the smaller sample--i.e., all adjusted FT/NFT differences are in the desired directions,

   -- indicate that FT/NFT differences are smaller in magnitude for the sample including big city schools, and

   -- adjusted differences range from .039 to .440 standard deviations of the dependent variables.

2. Sponsor level FT/NFT contrasts produce differences that are also similar to those based on the smaller sample--i.e., there is a great diversity of effects among different sponsors on different outcome measures. The particular contrast differences are presented in the table on the following page (A-63).

Adjusted FT/NFT Differences Based on Pooled School Means for Respective Sponsors *

	OUTCOME MEASURES							
		MAT				Locus of Control		
Sponsors (no.FT/no. NFT schools)	(1) WRAT	(2) Listening	(3) Reading	(4) Numbers	(5) Gumpgookies	(6) Posit.	(7) Neg.	(8) Absence
5. Bank St. College (16/15)	-.107	-.669	-.420	-.450	.598	.034	.165	-.156
7. U. of Oregon (11/10)	.673	1.397	.618	1.066	.045	-.0005	.554	-.140
8. U. of Kansas (20/15)	1.057	.918	.718	1.396	.793	-.495	-.202	-.036
9. High/Scope (11/12)	.438	.267	.706	.463	.567	.444	.188	-.414
10. U. of Florida (17/9)	.170	.148	.942	.050	.763	.232	.745	.121
11. EDC (13/12)	-.232	-.342	-.370	-.140	.541	-.073	-.039	-.375
14. SEDL (9/8)	-.091	-.320	.059	-.051	-.816	.250	-.020	-.903

---

*Taken from Figures VII-14 to VII-20. Differences are expressed in standard deviations of each variable. Note that three sponsors had no big city schools: 2. Far West, 3. U. of Arizona, 12. U. of Pittsburgh. Abt does not include these three in the sponsor-level analysis based on big city schools.

II. SPONSOR VIGNETTES: WITHIN SPONSOR, ONE-YEAR EFFECTS (Vol. IA, pp. VII-53 to VII-132)

The results summarized here are the product of analyses parallel to the sponsor level analyses reported in part one above. The present analyses examine individual sponsor effectiveness in promoting academic outcomes for children. There is one important difference. The analyses underlying the results reported here analyze sponsor performance three ways, using the classroom and the child as well as the school as the units of analysis.

A. Research Questions:

1. "What are the particular school effects of particular Sponsors?" (p. VII-56).

2. "What are the particular class level effects of particular Sponsors?" (p. VII-57).

3. "What are the particular child level effects of particular Sponsors?" (p. VII-58).

B. Sample: Three different samples were used.

1. Based on schools. This sample appears to be the same sample, including big city schools, used for the analyses of kindergarten effects described in part one.

2. Based on classrooms.

3. Based on children.

It is unclear how these samples were constructed. The report states: "Class and child level studies were based on a subset of classes and children included in the analytic group of schools. A somewhat different set of inclusion criteria were applied to classes and children for those studies, so that the subsets produced at class and child levels are different from the school level groups both in numbers and in characteristics" (p. VII-53).

The distribution of schools, classrooms, and children across sponsors is laid out in the table on the following page (A-65).

C. Variables:

1. Independent: exposure to one year of kindergarten experience with a given sponsor versus one year of experience in a NFT kindergarten.

2. Dependent: These analyses use the same eight outcome measures used in the "One-year, Kindergarten Outcomes"

Sample Size for Analyses Using Three Different Units of Analysis *

		School		Classroom		Child	
		FT	NFT	FT	NFT	FT	NFT
2.	Far West	29	20	37	29	302	303
3.	U. of Arizona	21	20	38	26	317	265
5.	Bank St. College	16	15	25	11	212	131
7.	U. of Oregon	11	10	21	12	153	119
*.	U. of Kansas	20	15	34	12	306	110
9.	High/Scope	11	12	25	17	173	144
10.	U. of Florida	17	9	23	9	243	101
11.	EDC	13	12	20	15	250	147
12.	U. of Pittsburgh	9	11	20	11	258	160
14.	SEDL	9	8	12	7	75	61
	Totals	156	132	255	149	2289	1541
		288		404		3830	

---

*Taken from Table III-2, p. III-6.

study summarized earlier: WRAT, MAT listening, MAT reading, MAT numbers, Gumpgookies, Locus of Control/positive, Locus of Control/negative, and absence. A ninth dependent variable is utilized in the analyses which use the child as the unit of analysis:

Peabody Picture Vocabulary Test (PPVT). "This is an individually administered measure of picture recognition vocabulary. The child is required to point to the correct picture corresponding to the word spoken by the examiner. The test was administered according to standard procedures; however, a number of pictures were modified to reflect Black rather than White ethnic characteristics. Thus, the test scores are not directly parallel to those derived from the standard version. Total scores equal the number of correct responses" (p. III-14).

3. Covariates: Different covariates were used for the three different analyses. The three sets are outlined in Table IV-1 which is taken directly from the Abt report.

D. Analytical Method: ANOVA and ANCOVA models were used to calculate unadjusted and adjusted FT/NFT differences. Nested designs were used to calculate respective sponsors' FT effects.

E. Comments: Readers are again referred to Chapter 3 of the main text for a discussion of the limitations imposed on the interpretation of findings by a range of methodological problems and shortcomings common to most of the FT evaluation efforts. Due caution should be exercised in dealing with the results of these sponsor vignettes. Issues worth special mention are the following:

1. Variables: As indicated in the summary of the One-Year Kindergarten Outcomes, although the reliability of most of the dependent variables are acceptable by conventional standards, the reliability of both locus of control measures is very low. Findings concerning these measures should be used with extreme caution.

   The fallibility of covariates remains unexamined. Interpretation of adjusted scores is consequently hindered by not knowing the quality of the adjustment.

   Perhaps the most opaque aspect of these analyses is the rationale for including some covariates (e.g., "standard deviation of Fall WRAT scores of tested children in the class"), and excluding others (e.g., "teacher's education" at the school level of analysis). In the absence of any discussion of the applicability of the covariates, these matters remain unclear.

Table IV-1. Overview of Variables Used as Covariates in Studies at Three Levels of Analysis

COVARIATE	PUPIL LEVEL OF ANALYSIS	CLASS LEVEL OF ANALYSIS	SCHOOL LEVEL OF ANALYSIS
1. FALL WRAT	Child's score	Mean score of (tested) children in the class	Mean score of (tested) children in the school
2. PRESCHOOL EXPERIENCE	Binary variable (any type of preschool experience vs. none)	N/A	N/A
3. MOTHER'S EDUCATION	Binary variable (high school or more vs. less than high school)	% of mothers of tested children in the class with a high school education or more	% of mothers of tested children in the school with a high school education or more
4. ADJUSTED INCOME INDEX	Parent's score on Adjusted Income Index	Mean score for tested children in the class	Mean score for tested children in the school
5. PARENTS' PERCEPTION OF SCHOOL'S RECEPTIVITY TO PARENT INVOLVEMENT	Parent's score on Receptivity Index	Mean score of parents of tested children in the class	Mean score of parents of tested children in the school
6. PARENT PARTICIPATION IN SCHOOL ACTIVITIES	Parent's score on Participation Index	Mean score of parents of tested children in the class	Mean score of parents of tested children in the school
7. YEARS AT CURRENT ADDRESS	Number of years family has lived at current address	Mean years at current address for tested children in the class	Mean years at current address for tested children in the school
8. TEACHER'S EDUCATION	Binary variable (bachelor's degree or higher vs. less than a bachelor's degree)	Binary variable (bachelor's degree or higher vs. less than a bachelor's degree)	N/A
9. TEACHER'S TEACHING EXPERIENCE	Number of years of teaching experience	Number of years of teaching experienc	N/A
10. ETHNICITY OF TEACHER	N/A	Binary variable (Black vs. non-Black)	N/A
11. FALL WRAT (STANDARD DEVIATION)	N/A	Standard deviation of Fall WRAT scores of tested children in the class	N/A
12. INTEGRATION	% of tested children in the class who are white	% of tested children in the class who are white	% of tested children in the school who are Black
13. % MINORITY	N/A	N/A	% of tested children in the school who are of a minority group
14. CITY SIZE	4 categories ($\geq$ 200,000; 50,000 $\leq$ 200,000; 10,000 $\leq$ 50,000; $<$10,000)	4 categories ($\geq$ 200,000; 50,000 $\leq$ 200,000; 10,000 $\leq$ 50,000; $<$10,000)	N/A
15. METROPOLITAN AREA	N/A	N/A	Binary variable indicating whether or not school is located in an SMSA
16. MIDDLE-SIZED CITY	N/A	N/A	Binary variable indicating whether or not school is located in a middle-sized city
17. WESTERN REGION	N/A	N/A	Binary variable indicating whether or not school is located in the Western Region
18. SOUTHERN REGION	N/A	N/A	Binary variable indicating whether or not school is located in the Southern Region

2. Analytical Method: Abt employs a "cross-validation strategy"--i.e., the use of three units of analysis--on the conviction that "where results are consistent for parallel questions across the three levels of aggregation, we have enhanced confidence that they represent the true effects." Haney* has demonstrated how this confidence is misplaced. There are at least three outside factors which influence the similarity and/or differences of scores across the different units of analysis. For one, overlapping but different samples were used for each of the three analyses. Differences might be due to sample differences; similarities might be due to the suppressing effect sample differences might exert on real differences. Two, different sets of covariates were used for the three analyses. Different sets of covariates may ". . . affect the effects' estimates at the different levels of aggregation."** And, three, the effects of aggregation itself need to be considered.

F. Results: The findings are summarized in the three tables which follow. Tables refer respectively to analyses based on the use of the school, the classroom, and the child as the unit of analysis. The entries in all three tables are the differences between the average of the respective FT units and the average of the respective NFT units, adjusted by using the indicated set of covariates, expressed in standard deviations of pooled scores on the respective measures. Positive values indicate that adjusted scores were higher for FT than for NFT. Positive differences are presumed preferable for all but the last outcome measure, absence, for which a negative difference is preferred.

The amount of information in the three tables is considerable. Reducing all of that information to manageable proportions is a formidable, if not impossible, task. Consequently, it is easy to understand the report's having resorted to merely providing a verbal description of the tables' contents. For our part, we focus on the child level analyses. We do so because it is our conviction that, if any meaning can be found in these results, it is to be found among the analyses which use the child as the unit of analysis. A few, very tenuous observations can be made at that level.

The results of the analyses which used the child as the unit of analysis indicate that there were five sponsors for whom

---

*Walt Haney, "Units of Analysis Issues in the Evaluation of Project Follow Through" (Cambridge, MA: Huron Institute, 1974), pp. 114ff.

**Haney, p. 118.

ADJUSTED FT/NFT DIFFERENCES BASED ON ANALYSES USING THE SCHOOL AS THE UNIT OF ANALYSIS

	Sponsors (No. FT/No. NFT schools)									
Outcome Measures	FW (29/20)	UA (12/20)	BC (12/20)	UO (11/9)	UK (16/12)	HS (8/9)	UF (15/9)	EDC (10/9)	UP (9/11)	SEDL (6/5)
WRAT	.023	.274	- .259	.756	1.289	.585	.316	- .031	.645	.510
MAT										
Listening	- .182	- .223	- .685	1.425	1.006	.453	.194	- .245	.0008	- .124
Reading	.283	.159	- .498	.604	.858	.842	.406	- .199	.113	.430
Numbers	- .064	- .463	- .495	.999	1.387	.634	.091	- .117	.969	.103
Gumpgookies	.607	.501	.631	- .015	.435	.664	1.080	.039	1.152	- .965
Locus of Control										
Positive	- .160	- .265	.392	- .055	- .276	.387	.233	.040	.314	.557
Negative	- .047	.160	- .120	.448	- .006	.092	.771	- .099	1.033	.152
Absence	.083	- .150	- .114	- .247	- .470	.781	.238	- .570	- .159	- 1.056

ADJUSTED FT/NFT DIFFERENCES BASED ON ANALYSES USING THE CLASSROOM AS THE UNIT OF ANALYSIS

	Sponsors (No. FT/No. NFT Classes)									
Outcome Measures	FW (37/29)	UA (38/26)	BC (25/11)	UO (21/12)	UK (34/12)	HS (25/17)	UF (23/9)	EDC (20/15)	UP (20/11)	SEDL (12/7)
WRAT	.008	- .205	- 3.480	.535	1.019	.229	1.454	- .231	1.468	.908
MAT										
Listening	.368	- .644	- 4.008	1.328	.413	.063	.348	.504	.518	2.205
Reading	.238	- .010	- 5.481	.885	.418	.294	.829	.163	.545	4.657
Numbers	.099	- 1.295	- 2.154	.586	1.264	.183	.539	.190	.658	.366
Gumpgookies	- .098	.212	- 2.661	- .101	.189	.016	1.406	.494	- .088	- 1.887
Locus of Control										
Positive	.087	- .686	- 4.168	.096	- .141	.830	.081	.019	.655	- .523
Negative	- .111	- .144	- 1.925	.540	- .156	1.028	1.357	.139	1.223	- .020
Absence	- .208	- .196	.945	- .587	.172	.032	- .179	1.402	- .350	- 2.245

ADJUSTED FT/NFT DIFFERENCES BASED ON ANALYSES USING CHILDREN AS THE UNITS OF ANALYSIS

	Sponsors (No. FT Children/No. NFT Children)									
Outcome Measures	FW (302/ 303)	UA (317/ 265)	BC (212/ 131)	UO (153/ 119)	UK (306/ 110)	HS (173/ 144)	UF (243/ 101)	EDC (250/ 147)	UP (258/ 160)	SEDL (75/ 61)
WRAT	.084	.028	.296	.550	.605	.198	.991	- .160	.751	.157
PPVT	.041	.031	.004	.015	- .157	.260	.796	- .073	.250	.346
MAT										
Listening	.112	- .417	- 1.008	.742	.156	.144	.380	.121	.373	.008
Reading	.128	- .013	- .567	.504	.296	.229	.656	.029	.310	1.027
Numbers	.012	- .496	.128	.713	.695	.178	.619	.108	.449	- .044
Gumpgookies	.012	.082	- .136	.029	.141	- .059	.376	- .111	.029	- .483
Locus of Control										
Positive	.097	- .147	.030	.142	.112	.237	.076	- .107	.227	.284
Negative	- .070	- .009	.459	.166	.037	.396	.374	.134	.568	.268
Absence	- .147	.159	- .593	.102	.150	- .146	- .159	.218	.354	- .235

there were one or less NFT-favoring differences. Differences for two sponsors were all FT-favoring, i.e., positive on all measures except absence. Differences for three sponsors--Far West, University of Oregon, and High/Scope--favored FT on all but one of the outcome measures.

Just how important these findings are remains very much in doubt. First, none of the differences is especially big. Second, it is dangerous to treat all outcome measures as equivalent. Third, the assumption that differences can be attributed exclusively or primarily to sponsor treatments cannot be made with much confidence.

III. PRESCHOOL EFFECTS (Vol. IA, pp. VIII-29 to VIII-41)

One of the questions behind the FT evaluation was whether FT experience maintained benefits allegedly produced by Head Start experience. This particular set of analyses comes closest (among those found in Abt's 1974 report) to addressing the issue of continuity of effect.

A. Research Questions:

1. Do children with HS experience, some other kind of preschool experience, or no preschool experience do best in FT?

2. Do different sponsors have systematically different effects on children with Head Start (HS), some other (OPS), or no preschool experience (NPS)?

B. Sample: The sample appears to be similar to the sample used in the analyses of the Sponsor Vignettes which employed the child as the unit of analysis. A comparison of the size of this sample (N = 3,580) with that used in the kindergarten child outcome analyses (N = 3,830) indicates, however, that they are not identical.

Abt mentions three inclusion rules for the formation of the present sample:

1. A complete set of data: Fall 1971 and Spring 1972 test scores for children as well as parent and teacher measures.

2. Limited to only Black and White children.

3. Eliminate all children who had both HS and OPS experience.

The distribution of children across sponsors in the resulting analytic subset is described in the following table (see page A-74).

C. Variables:

1. Independent:

   a. FT or NFT experience during kindergarten
   b. Association with one of ten sponsors
   c. Head Start, some other, or no preschool experience

2. Dependent: The same set of outcome measures used for the child level analysis of kindergarten outcomes is used here:

Preschool Experience		Sponsors									
		2	3	5	7	8	9	10	11	12	14
FT (N=2087)	HS (N=1166)	177	131	95	111	180	93	88	106	138	47
	Other Preschool (N=336)	36	52	29	10	29	21	53	65	37	4
	No Preschool (N=585)	47	101	79	13	72	39	53	65	7[illegible]	21
NFT (N=1493)	HS (N=329)	83	50	9	33	53	30	10	13	34	14
	Other Preschool (N=339)	57	88	28	27	13	32	23	29	27	5
	No Preschool (N=825)	143	118	94	59	41	79	68	88	93	42

85

a. Wide Range Achievement Test
b. Peabody Pictures Vocabulary Test
c. Metropolitan Achievement Test-Reading
d. MAT-Arithmetic
e. MAT-Listening for Sounds
f. Gumpgookies
g. Locus of Control-positive
h. Locus of Control-negative
i. Absence

3. Covariates: The same set of covariates used for the child level analysis of kindergarten outcomes is used here with the important exception that preschool experience, a covariate in the former child level analyses, has become one of the independent variables in the present analysis. The nine covariates are (p. VIII-34):

   a. Fall WRAT. This test served as the pretest covariate for all measures except the PPVT, which used the Fall PPVT as the pretest covariate.

   b. Adjusted income. A measure reflecting family income adjusted for family size and urban/rural residence.

   c. Mother's education. Categorized as high school diploma or greater versus less than high school diploma.

   d. Years at current address.

   e. School receptivity. As perceived by the parents.

   f. Parent-school involvement. As reported by the parents.

   g. Teacher's years of education.

   h. Teacher's years of teaching experience.

   i. Percentage of White pupils in the classroom.

   j. City size. Coded on a scale of one through four for populations ranging from under 10,000 to over 200,000.

D. Analytical Method: A "fully crossed analysis of covariance" was used. The child serves as the unit of analysis. The fallibility of Fall WRAT scores was apparently taken into account and adjustments made, although no explanation is provided as to how these adjustments were actually made.

E. Comments: As with all analyses summarized in this appendix, we refer readers to the general remarks in Chapter 3 concerning the logical and methodological difficulties encountered by all the FT evaluation analyses. We restrict our remarks here to those which pertain specifically to this set of analyses.

1. Significant preschool-by-sponsor-by-FT/NFT interaction was detected on WRAT, MAT-reading, MAT-math, and MAT-listening analyses. Similarly, significant sponsor-by-FT/NFT interaction was detected for WRAT, MAT-reading, MAT-math, MAT-listening, and PPVT test scores. Main effects' estimates based on aggregations of sponsors or preschool experience are appropriately flagged as possibly misleading and the analysis turns to within-sponsor, intergroup differences. However, when this subsequent analysis is taken up, only three outcome measures are dealt with. No explanation is given concerning the reduction of the complete set of academic measures to the three used.

2. The graphic presentations of FT/NFT differences for each of the three preschool groups (Vol. IA, pp. VIII-37 to VIII-39) is confusing if not misleading. The units on the vertical axes are not defined. The horizontal axis is not an axis in the normal sense of the word since FT/NFT is obviously nominal, not an interval or even ordinal, measure. The slopes of the respective group lines indicate--but only in an undefined way--the differences between FT and NFT children.

F. Results:

1. The analytical model calls for the examination of three main effects.

   a. FT/NFT adjusted contrasts indicated no statistically significant differences on MAT-listening, PPVT, Gumpgookies, Locus of Control positive or negative, or absence. Three statistically significant ($p < .005$) differences were obtained on WRAT, MAT-reading, and MAT-math. However, although significant, the FT/NFT contrast was able to explain less than one-half of one percent of the variation of any of these three outcome measures.

   b. Sponsor contrasts were found to be confounded by interactions with FT/NFT and preschool experience. No results on this second main effect are reported by Abt.

   c. Preschool experience contrasts produced the following results (see page 80).

2. Analysis of possible interaction effects, however, detected significant (1) Sponsor-by-FT/NFT and (2) Preschool experience-by-sponsor-by-FT/NFT interactions on most of the academic outcome measures (see page 81).*

---

*Preschool experience-by-FT/NFT interaction was not significant in the analyses of any of the nine outcome variables.

Outcome Variable*	Amount of Variance Attributable To Type of Preschool Experience	Level of Significance
WRAT	.0592	.005
MAT		
Reading	.0304	.005
Math	.0305	.005
Listening	.0277	.005
PPVT	.0891	.001
Gumpgookies	.0080	.005
Locus of Control		
Positive	.0096	.005
Negative	.0016	nss
Absence	.0001	nss

The amount of variance attributable to differences in preschool experience (HS, OPS, NPS) is not significant on two outcome measures. While it is significant on seven, the amount of variance attributable to preschool experience difference is less than one percent in the case of Gumpgookies and Locus of Control-positive. This leaves the five academic outcome measures. The amount of variance attributable to preschool experience differences ranges from 2.8% for MAT-listening to 8.9% of PPVT.

---

*Taken from Table A VIII-1, Volume IB, p. V-7.

Outcome Variable*	Interactions Sponsor X FT/NFT	 Preschool X Sponsor X FT/NFT
WRAT	.005	.05
MAT		
Reading	.005	.05
Math	.005	.01
Listening	.005	.10
PPVT	.01	nss
Gumpgookies	nss	nss
Locus of Control		
Positive	nss	nss
Negative	nss	nss
Absence	nss	nss

Abt concludes that "the significant interactions suggest that in the achievement areas FT/NFT differences vary across Sponsors and among the three different categories of preschool experience. Therefore, child-level analyses which combine the three preschool groups (such as the Sponsor by FT/NFT interaction) or combine the Sponsors (such as the preschool by FT/NFT interactions) are inappropriate and misleading" (p. VIII-36, our emphasis).

---

*Taken from Table A VIII-1, Volume IB, pp. V-7 to V-8. Table entries are significance levels of respective interactions on each dependent variable.

3. Since comparisons which grouped children across sponsor or across type of preschool experience were deemed inappropriate, the analysis turned to within-sponsor, preschool group differences on adjusted spring scores for three academic outcome measures: WRAT, MAT-math, and MAT-reading. This approach amounted to 90 separate analyses--three outcome measures times ten sponsors times three preschool groups. The results are extremely varied and thus difficult to summarize.* We reproduce here Abt's findings without the report's commentary. These "graphs" are presented with great reservation since, as discussed in "Comments" above, this visual device can be very misleading. We leave it to the reader to interpret the results.

---

*Abt attempts some generalizations which, when examined closely, prove to be unwarranted, e.g., "Some Sponsors appear to produce roughly equivalent effects in all three preschool groups, for example the University of Arozona (3) and the University of Oregon (7) on all three measures . . ." (p. VIII-36). "Roughly equivalent effects" is a somewhat ambiguous term, but even by the wildest stretch of common English terminology, it is difficult to see how it could refer to the dissimilar results obtained between these two sponsors. For example, while University of Arizona FT children in all three preschool groups score as low as or lower than NFT children, University of Oregon FT children in all three groups score as high as or higher than NFT children.

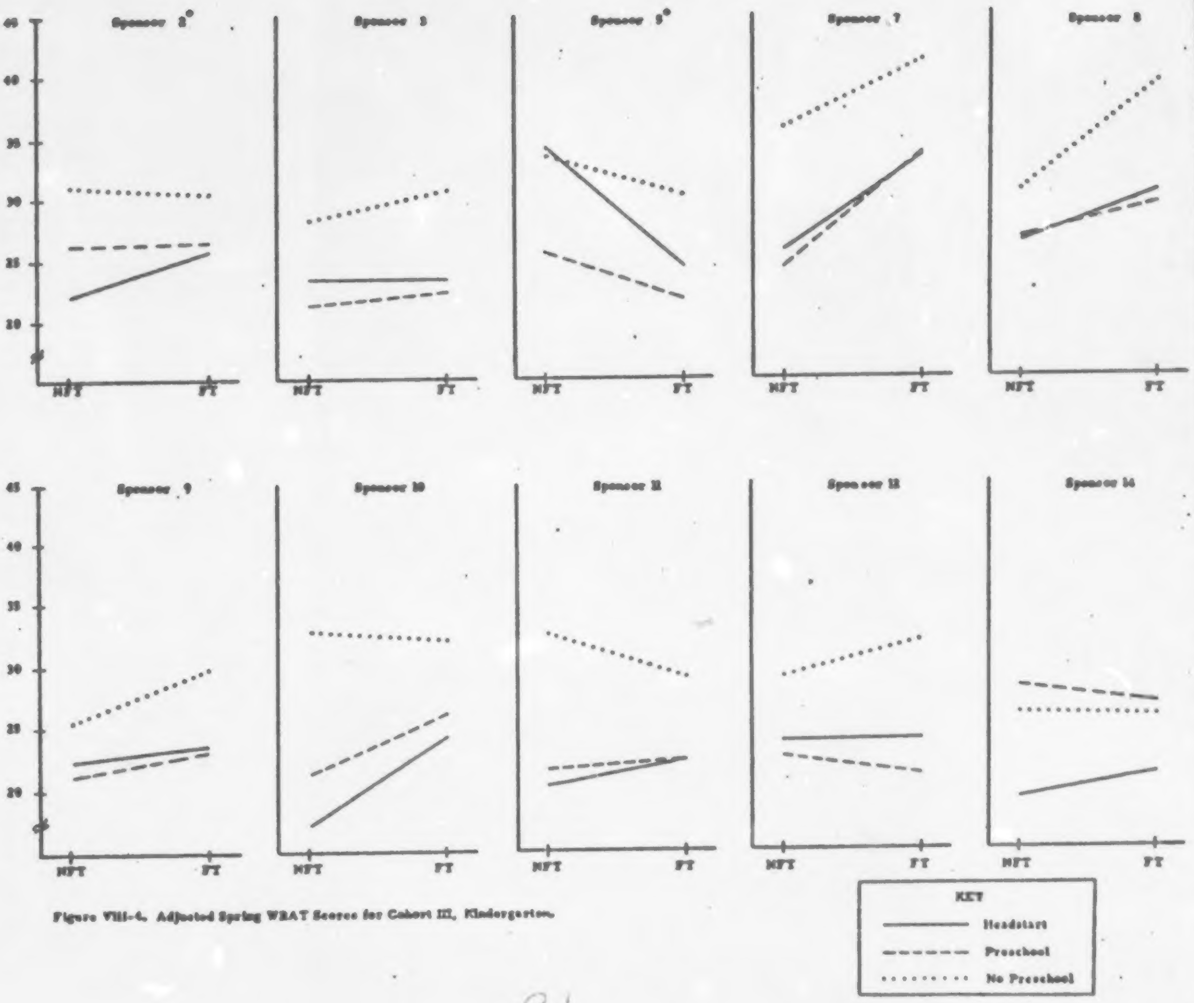


Figure VIII-4. Adjusted Spring WRAT Scores for Cohort III, Kindergarten.

91

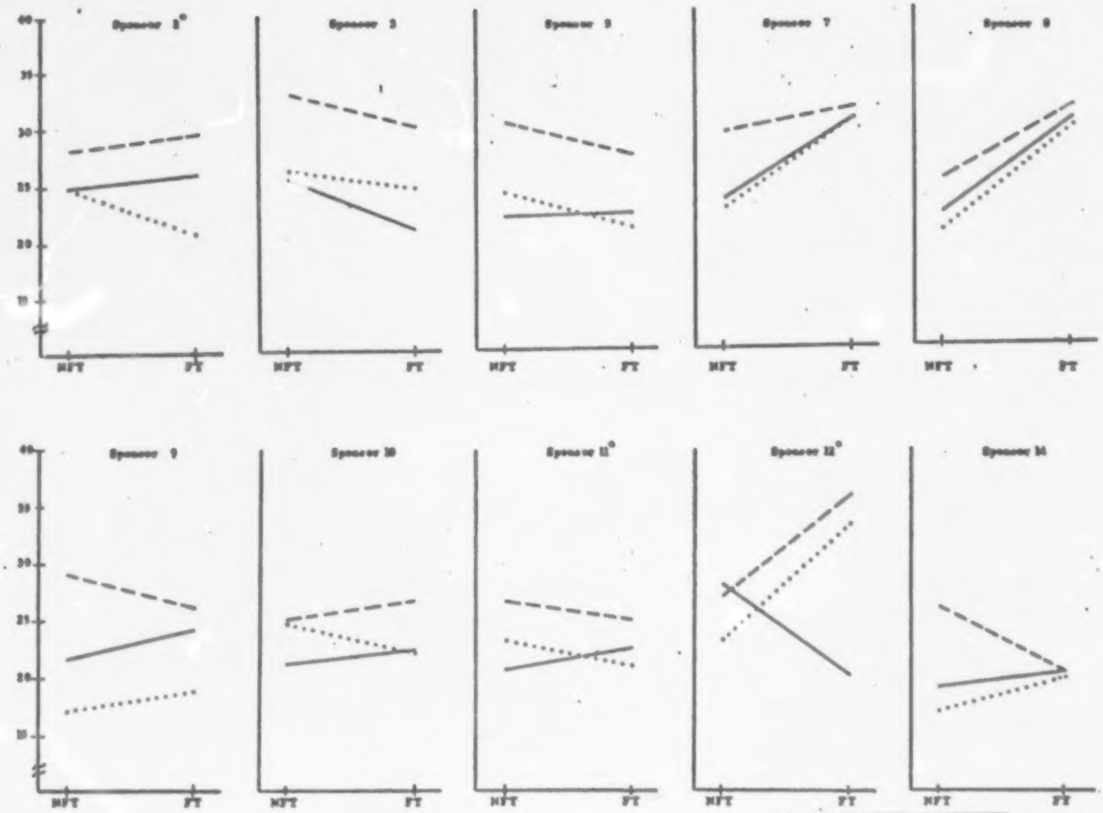


Figure VIII-5. Adjusted Spring MAT-Math Scores for Cohort III, Kindergarten.

KEY

——— Headstart

– – – – Preschool

· · · · · · No Preschool

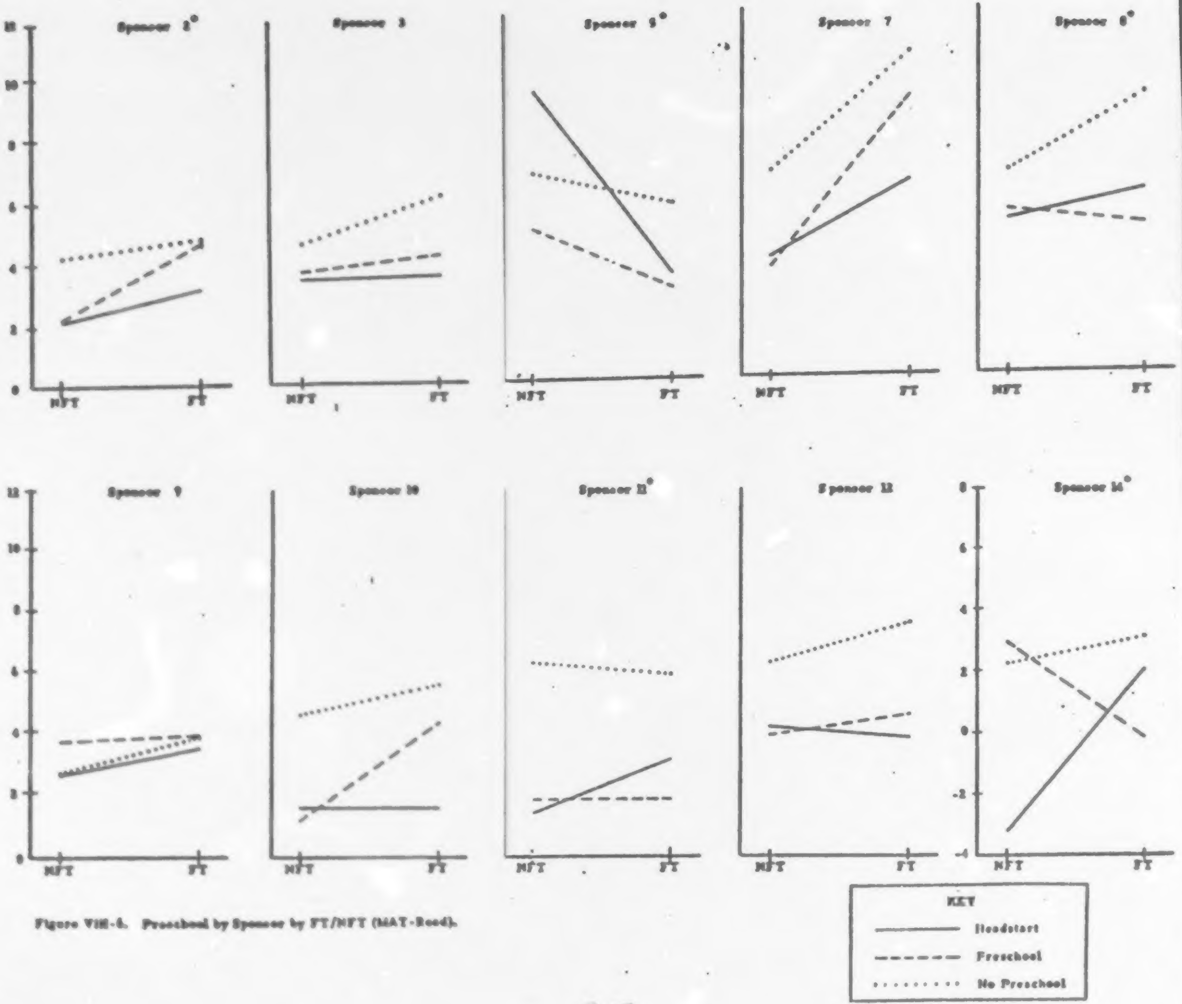


Figure VIII-5. Preschool by Sponsor by FT/NFT (MAT-Read).

IV. CHILD SEX DIFFERENCES (Vol. IA, pp. VIII-54 to VIII-61)

A. Research Questions:

1. Do boys and girls differ on outcome measurements?

2. Do some sponsors have effects on boys different from those on girls and, if so, on what outcomes?

B. Sample: The sample is a sample of children, N = 3,830. It appears to be the same sample as that used for the child level analyses in the Sponsor Vignettes (see section II above).

C. Variables:

1. Independent:

   a. Sex of the child
   b. Sponsor program
   c. FT/NFT experience

2. Dependent: Nine outcome measures were used in this analysis:

   a. Spring WRAT
   b. PPVT
   c. MAT-reading
   d. MAT-math
   e. MAT-listening
   f. Gumpgookies
   g. Locus of control-positive
   h. Locus of control-negative
   i. Absence

3. Covariates: The following twelve covariates were used.

   a. Fall WRAT
   b. Fall PPVT (for PPVT outcome only)
   c. Preschool experience
   d. Adjusted income
   e. Mother's education
   f. Years at current address
   g. School receptivity
   h. Parent-school involvement
   i. Teacher's years of education
   j. Teacher's years of experience
   k. Percentage of white pupils in a classroom
   l. City size

D. Analytical Method: A three-factor, fully crossed ANCOVA was used to determine the significance of both the sex main effect and the interaction of sex-by-sponsor-by-FT/NFT. The child was the unit of analysis.

E. Comments: No specific comments are offered concerning this analysis. Readers are referred to Chapter 3 of the main text for general comments on these and other FT evaluation analyses.

F. Results:

1. Significant sex differences were detected on WRAT, MAT-numbers, PPVT, Gumpgookies, and Locus of Control-positive. However, although the differences were significant, none approach .25 standard deviations in magnitude. For whatever they are worth, the particular differences were:
    a. Boys outscored girls on MAT-numbers (.12 s.d. units), PPVT (.06 s.d. units), and Locus of Control-positive (.04 s.d. units).
    b. Girls outscored boys on WRAT (.06 s.d. units) and Gumpgookies (.07 s.d. units).
2. No significant sex-by-sponsor-by-FT/NFT interaction was detected on any of the nine outcomes. It therefore appears that respective sponsors do not have systematically differentiated effects for boys and girls.

V. CHILD ETHNICITY (Vol. IA, pp. VIII-42 to VIII-53)

A. Research Question: Do respective sponsors have systematically different effects on either Black or White children?

B. Sample: The sample is a sample of children (N = 3,830). It appears to be the same as that used for the child level analyses in the Sponsor Vignettes described earlier. For a description of students across sponsors, see section II above. In terms of ethnic groups, the children were distributed as follows:

	Black	White	Total
FT	1,217	1,070	2,287
NFT	507	1,034	1,541
Total	1,724	2,104	3,828

C. Variables:

1. Independent:
    a. Ethnicity of child--Black or White
    b. Sponsor program
    c. FT/NFT experience

2. Dependent: The same nine outcome measures as were used in analyses already described were used in this analysis.

3. Covariates: The following twelve covariates were used:
    a. Fall WRAT
    b. PPVT (for PPVT outcome only)
    c. Preschool experience
    d. Adjusted income
    e. Mother's education
    f. Years at current address
    g. School receptivity
    h. Parents' involvement in school
    i. Teacher's education
    j. Teacher's experience
    k. Percentage of Whites in classroom
    l. City size

D. Analytical Method: A three-factor, fully crossed ANCOVA was used. Focus was placed on sponsor-by-FT/NFT-by-ethnicity interaction.

E. Comments: Some of the same criticisms raised concerning the presentation of results in the Preschool Experience Study described earlier apply here. Specifically, the graphical plots of FT/NFT contrasts for Blacks and Whites for each sponsor are likely to be misleading. The scale on the vertical axis is not defined and the horizontal "axis" is not an axis in the ordinary sense of the term. Slopes indicating differences between FT and NFT children within each ethnic group are as much a function of the arbitrary distance selected for FT and NFT on the horizontal "axis" as they are of real differences.

F. Results:

1. Significant ethnicity-by-sponsor-by-FT/NFT detected on four outcomes: WRAT, MAT-numbers, MAT-listening, and PPVT. This interaction suggests ". . . that Black children and White children do not respond in the same way to all Sponsor's FT programs" (p. VIII-46).

2. "Significant interaction also indicates that for those achievement outcomes it is inappropriate to combine different groups of children across either Sponsor or ethnicity to look at two-way interactions" (p. VIII-46, emphasis in original).

3. No significant ethnicity-by-sponsor-by-FT/NFT interaction was detected for motivational measures or absence.

4. Given the apparent inappropriateness of pooling children across ethnic groups, FT/NFT differences on four academic outcomes for both Blacks and Whites were examined at the sponsor level. The results indicate that children of one racial group do not do differentially better or worse than children of the other racial group in most models.

VI. ENTERING CLASS-ACHIEVEMENT LEVEL (Vol. IA, pp. VIII-13 to VIII-28)

A. Research Questions:

1. The substantive question was "Do some Sponsors have systematically different effects on classes that start out at different levels of achievement?"

2. A methodological question was also addressed: "Are the relationships between the initial achievement level and posttest performance in a Sponsor's FT and NFT classes sufficiently similar to justify adjustment for initial difference and comparison?" (p. VIII-13).

B. Sample: The sample is a sample of classes and appears to be the same as that used for the classroom analyses in the Sponsor Vignettes described above.

C. Variables:

1. Independent:

   a. Entering achievement level of the class expressed as the class mean for Fall, 1971 WRAT scores
   b. Sponsor program
   c. FT/NFT experience

2. Dependent: The familiar set of nine outcome measures were used here--i.e., Spring WRAT, MAT-reading, MAT-arithmetic, MAT-listening, Gumpgookies, Locus of Control-positive and negative,* and absence. All measures were expressed in class means.

3. Covariates: Class means for the following ten variables were used:

   a. Mother's education
   b. Years at current address
   c. Parent perceived receptivity of the school
   d. Parent participation
   e. Adjusted income index
   f. City size
   g. Teacher's ethnicity
   h. Teacher's education
   i. Teacher's experience
   j. Percent White in class

D. Analytical Method: Three factor, crossed ANOVA and ANCOVA models were used to determine the significance of class entry level-by-sponsor-by-FT/NFT interaction.

---

*We presume that both positive and negative scores were used. The report does not explicitly mention both (p. VIII-14).

E. Comments: Although the analysis of three-way interactions yielded significant results in the cases of Spring WRAT ($p < .005$), MAT-listening ($p < .05$), and MAT-math ($p < .10$), the explanation of homogeneity of regression at the sponsor level focuses on Spring WRAT and MAT-reading. Two questions arise as a result: (a) Why was homogeneity of regression examined in reference to a variable (MAT-reading) on which no significant interaction was detected? (b) Why was homogeneity of regression not examined in reference to two other variables on which significant interactions had been detected?

In the first case, finding no heterogeneity of regression with reference to MAT-reading is hardly surprising. What is surprising is that homogeneity of regression was examined in reference to this variable at all. In the second case, since the homogeneity of regression of MAT-listening and MAT-math on Fall WRAT is unexamined and undiscussed, its presence or absence remains unknown.

The set of covariates is not exactly the same as that used in the class level analyses of the Sponsor Vignettes nor for the Class Ethnic Composition analysis described below (another analysis using the classroom as the unit of analysis). The differences among the sets of covariates used for the three analyses are:

1. The class level analyses of the Sponsor Vignettes purportedly use all twelve covariates mentioned in section II above, while this analysis uses neither Fall WRAT scores (for understandable reasons) nor Teacher's Ethnicity. The latter deletion is not explained.

2. The Class Ethnic Composition analysis described below deletes percentage White pupils in the class (for understandable reasons) and the standard deviation of Fall WRAT scores. The uncommented upon exclusion of this covariate is as inexplicable as its inclusion was in the first place.

F. Results:

1. Significant three-way interactions (indicating differential sponsor effects for different entry level groups) were found in some cases. Using the ANOVA model, significant interaction ($p < .005$) was detected only in the case of Spring WRAT scores. On the basis of the ANCOVA, significant interactions were identified in the case of MAT-listening ($p < .05$) and MAT-math ($p < .10$), as well as Spring WRAT ($p < .005$). In general, it appears that some of the sponsors have different effects on different entry level groups. Abt notes the following particular cases of differential effectiveness.*

---

*Unmentioned sponsors did not appear to have differential effects.

a. Far West Laboratory's ". . . program is apparently more effective with classes of higher-than-average initial achievement and less effective with classes of lower initial achievement" (p. VIII-20).

b. University of Oregon's figures ". . . indicate substantially greater gains for higher initial achievement level in FT classes" (p. VIII-24).

2. The presence of significant three-way interaction on some outcome measures suggested heterogeneity of regression relations. The analysis turned to an examination of the homogeneity of regression between FT and NFT groups at the sponsor level. Two outcome measures were used--Spring WRAT and MAT-reading.

"The results indicate heterogeneity of regression only for the Spring WRAT and only for some Sponsors. Furthermore, for two of the four Sponsors for whom heterogeneity is found, the heterogeneity can be accounted for in terms of the distribution of the sample of classes across initial achievement level in the Sponsor NFT group. For Sponsors 5 and 10, the NFT groups show regressions of Spring WRAT on Fall WRAT that are substantially different from their FT groups as well as from all other groups in this study. In both cases, the importance of the disparate regression is brought into question by the fact that these groups contain a very restricted representation across initial achievement level. For these Sponsors the NFT regression is highly unstable. The instability of the regression suggests that the heterogeneity is spurious. For the other two Sponsors, Sponsors 2 and 7, the heterogeneity is obtained because the FT group has a substantially steeper regression than the NFT group, which displays a regression that is very similar to other NFT groups in the sample. For both Sponsors the steeper slope in FT indicates a greater gain as initial achievement level increases. For Sponsor 2, the FT classes with higher-than-average initial achievement level gain more than NFT classes with a comparable initial achievement level. Classes of average or below average achievement level do not fare as well. For classes of average initial achievement level FT and NFT produce equal outcome scores and for classes of lower-than-average initial achievement level, the NFT classes exceed the FT classes. For Sponsor 7, a substantially different picture is obtained. Sponsor 7's FT classes exceed the NFT classes across all levels of initial achievement score. However, the advantage of FT increases with an increase in initial achievement level.

"Possibly the most important aspect of the results is the general inadequacy of the sample size for assessing the relationship between initial achievement level and outcomes. For most Sponsors, the number of classes in FT and NFT is minimally appropriate for the estimation of regression effects and much of the heterogeneity can be accounted for in terms of the inappropriateness of the distributions of classes across initial achievement level" (pp. VIII-27 to VIII-28).

3. One of the most important findings reported in this analysis appears in a footnote: "All other covariables [i.e., all except Fall WRAT] were found to be homogeneous across Sponsors' FT and NFT groups" (p. VIII-13).

VII. CLASS ETHNIC COMPOSITION (Vol. IA, pp. VIII-1 to VIII-12)

A. Research Questions:

1. Do different sponsors "have different effects on classes which are integrated to different degrees?" (p. VIII-1).

2. "Does ethnic composition of the class relate to class performance?" (p. VIII-1). This latter question is operationalized and expressed in two hypotheses:

    a. Outcome measures are a linear function of the percentage of White students in a class.

    b. Outcome measures are a non-linear function of the percentage of White students in a class.

B. Sample: A sample of classes is used, N = 404. The sample appears to be the same as that used for the class level analyses in the Sponsor Vignettes described earlier.

C. Variables:

1. Independent:

    a. Integration of the class is expressed in terms of:

        (1) the proportion of White students in the classroom (as a linear term) and
        (2) the square root of the proportion of White students in a class (as a curvilinear term).

    b. Association with one of ten sponsors

    c. FT or NFT experience

2. Dependent: Spring class means for:

    a. WRAT
    b. MAT-reading
    c. MAT-arithmetic
    d. MAT-listening
    e. Gumpgookies
    f. Locus of control-positive
    g. Locus of control-negative
    h. Absence

3. Covariates: Class means for:

    a. Fall WRAT
    b. Years at current address
    c. Adjusted income

d. Teacher's experience
e. Parent's perceptions of the receptivity of the school
f. Parent participation
g. Percentage of mothers with a high school education
h. Teacher's education
i. Teacher's ethnicity
j. City size

D. Analytical Method: A "multiple regression analogue of ANCOVA" with four terms was used: (1) a set of covariates, (2) "an effects coded variable for FT/NFT membership," (3) "a variable representing proportion white in class [linear component]," and (4) "a variable representing the square root of proportion white in class [non-linear component]" (pp. VIII-3 to VIII-4).

E. Comments: The reader is referred to general comments in Chapter 3.

F. Results:

1. Although not raised as a principal research question, the analysis produced results concerning the main effect of class ethnic composition. The proportion of variance in outcome measures attributable to the linear term used to express class ethnic composition was found to be significant only in the case of Gumpgookies ($p < .10$) and Locus of Control-positive ($p < .005$). The amount of variance attributable to the combined linear and non-linear terms was significant in the case of Gumpgookies ($p < .05$), Locus of Control-positive ($p < .005$), and absence ($p < .025$). The report cautions: "We, however, are aware that the low reliability of those affective measures may be contributing some interesting but spurious results" (p. VIII-12).

2. "The distribution of classes across five categories of ethnic composition is presented in Table VIII-2 for pooled FT and NFT groups and for each Sponsor's FT and NFT groups. The table indicates that the pooled FT and NFT groups are fairly comparable in their distribution of classes across categories of proportion white in class. However, the FT group has approximately three times as many classes as the NFT group in the predominantly non-white category and this ratio of 3 to 1 is repeated in many of the individual Sponsors' distributions" (p. VIII-6).

3. Concerning the assessment of within-sponsor class ethnicity effects:

   "For most Sponsors, the ethnicity distribution of classes is far two skewed to permit a meaningful analysis. Sponsors 5, 10, 11, and 12 have fewer than a quarter of their classes in the middle ethnicity range [proportion White > 20% and ≥ 80%]. Sponsors 7, 8, 9, and 14 have a very small percentage of their classes in the predominantly White [> 80%] category. Both the small numbers of observations within these Sponsors' groups as well as their skewed ethnicity distributions obviate the assessment of ethnic composition effects within these Sponsors.

"The remaining Sponsors, 2 and 3, have predominantly nonwhite classes in their FT groups with their other classes spread more or less evenly across the range of proportion White in class. Although these Sponsors have some spread in their distribution of FT classes across proportion White in class, the number of classes in each category is quite small" (p. VIII-11).

4. Concerning the issue of the shape of the relationship between class ethnic composition and outcome measures:

   "The linear hypothesis that outcomes change uniformly with changes in the ethnic composition of the class is supported only for the Locus of Control for positive events. The regression coefficient for proportion White in class and class Locus of Control for positive events was .17, indicating a weak positive relationship: Classes with a higher concentration of White pupils also exhibit higher scores [more internal] on positive Locus of Control.

   "The non-linear hypothesis fits the data as badly as the linear hypothesis for the achievement outcomes. However, for two of the affective measures and the Absence outcome, the non-linear hypothesis is supported" (p. VIII-6).

   In general:

   "At best these results and this study must be looked at as exploratory on the one hand and inconclusive on the other. the sample does not generally permit the assessment of effect by Sponsors; therefore, biases may be present for which appropriate adjustments cannot be made. Further, the assessment of the overall integration effect is tenuous because of the pooling of potentially confounding factors and the disproportionality of the distribution. However, the effects on the Gumpgookies, Locus of Control, and Absence outcomes are anticipated by previous theoretical work and the results provide a promising background for future research" (p. VIII-12).

TABLE VIII - 2

Number of Classes Falling into Each of Five Categories of Percent White in Class for Each Sponsor's FT and NFT group and for The Pooled FT and NFT Groups (p. VIII-7).

Sponsor	< 20%	> 20% < 40%	> 40% < 60%	> 60% < 80%	> 80%
2 FT	18	16	4	4	5
NFT	11	1	3	1	13
3 FT	13	2	5	9	9
NFT	4	3	1	5	13
5 FT	8	0	1	4	12
NFT	0	0	2	1	8
7 FT	15	2	1	2	1
NFT	7	1	0	2	2
8 FT	30	0	1	3	0
NFT	9	0	1	2	0
9 FT	18	6	1	0	0
NFT	6	6	0	1	4
10 FT	14	0	2	2	5
NFT	3	0	2	1	3
11 FT	12	1	0	0	7
NFT	7	0	1	0	7
12 FT	6	0	0	0	14
NFT	1	2	0	0	8
14 FT	6	1	5	0	0
NFT	0	3	3	1	0
TOTAL FT	126	28	15	33	53
NFT	48	16	13	13	58
TOTAL	174	44	28	46	111

VIII. COHORT I THREE-YEAR EFFECTS (Vol. I A, pp. VII-133 to VII-144)

A. Research Question: Do three years of FT experience have a demonstrable effect?

B. Sample: Just how the sample used in this analysis was constructed is not made clear. It was composed of 1,216 students all from Cohort I. These students had entered first grade in the fall 1969 and were tested then and during the spring of their third grade (1972). All had had Head Start experience prior to first grade. The 40 schools they attended were distributed across six sponsors as follows:

	6 Sponsors:	5	6	7	9	11	12	
40 Schools	FT:	4	8	3	6	1	3	25
	NFT:	2	3	2	3	2	3	15

C. Variables:

1. Independent:

    a. FT or NFT experience;

    b. Affiliation with one of six sponsors.

2. Dependent--spring 1972 scores on:

    a. MAT-reading;

    b. MAT-arithmetic;

    c. MAT-spelling.

3. Covariates: Since "no psychometrically-equivalent pretest was administered in 1969 at the beginning of the first grade, we therefore used the results of the Pre-School Inventory (PSI) and Wide Range Achievement Test (WRAT) as [surrogate] measures of entry-level achievement for purposes of covariate adjustment." (p. VII-134)

D. Analytical Method: ANOVA and ANCOVA -- presumably crossed models. The school served as the unit of analysis.

E. Comments: In addition to the general comments of Chapter 3, the following points should be noted.

1. The validity and reliability of the dependent variables and the covariates are neither discussed nor reported.

2. It is unclear from the report whether within-sponsor, FT/NFT contrasts are adjusted or unadjusted.

F. Results:

1. No significant ($p \leq .05$)NFT/NFT main effect was detected on any of the three outcome measures.

2. A significant ($p \leq .05$) sponsor effect was detected on the spelling outcome -- i.e., children associated with different sponsors scored differently.

3. Significant ($p \leq .05$) sponsor-by-FT interaction was observed for both the reading and spelling outcome measures.

4. Within-sponsor, FT/NFT contrasts turned up fairly consistent patterns:

   a. FT children in Bank Street College, University of Georgia, EDC and University of Pittsburgh programs scored as well as or better than NFT children on all three outcome measures.

   b. FT children in University of Oregon and High/Scope programs scored lower than NFT children on all three outcomes.

5. The covariates taken together account for approximately one-half the variance in each of the three outcomes.

IX. INTER-COHORT ANALYSIS (Vol. I A, pp. VII-144 to VII-153)

A. Research Questions: Two sets of questions were addressed.

1. The first set had to do with sponsor effects (see pp. VII-144 - VII-146):

   a. Are there detectable sponsor differences across cohorts?

   b. Are there FT/NFT differences across cohorts?

   c. Is there any sponsor-by-FT/NFT interaction across cohorts?

2. The second set had to do with differences between Cohort I and Cohort II children:

   a. Are there detectable differences between cohorts across sponsors?

   b. Is there any sponsor-by-cohort interaction?

   c. Is there any FT/NFT-by-cohort interaction?

   d. Is there any sponsor-by-cohort-by-FT/NFT interaction?

B. Sample: The sample is a sample of schools. It appears to be an opportunity sample based on the availability of data. It was limited to 24 schools which had both Cohort I EK and Cohort III EK children and had had the same sponsor for both cohorts. The 24 schools were distributed across sponsors as follows:

Sponsors:	6 Sponsors	:	2	3	5	8	11	13	
Sites:	24 Schools	FT:	3	2	3	3	2	2	15
		NFT:	2	1	1	1	2	2	9
									24

C. Variables: As the research questions imply, there were two analyses. Although the same dependent variable and covariate were used for both, each had its own independent variable(s).

1. Independent:

   a. Analysis A (across cohorts):

      -- Sponsor affiliation;

      -- FT or NFT experience.

b. Analysis B (across sponsors):

-- Cohort membership.

2. Dependent: Both analyses used end of kindergarten WRAT scores.

3. Covariates: Both analyses used entering WRAT scores.

D. Analytical Methods: Two ANCOVA's were undertaken using a multiple regression format. The respective models were:

1. Spring WRAT Scores = Fall WRAT Scores + Sponsor affiliation + FT/NFT experience + Sponsor by FT/NFT interaction .

2. Spring WRAT Scores = Fall WRAT Scores + Cohort membership + Sponsor by Cohort interaction + FT/NFT by Cohort interaction + Sponsor by FT/NFT by Cohort Interaction

The school is the unit of analysis.

E. Comments: The report indicates that the present analyses are "by way of preparing for better data to come...." (p. VII-144). In light of this intent and the many shortcomings of the analyses -- e.g., no examination of representativeness of the sample, no determination of the fallibility of any of the measures -- we suggest that the analyses be taken as methodological exercises and not as the source of valid results concerning substantive issues.

F. Results:

1. Analysis A.

a. Sponsor effects across cohorts were found to be significantly ($p \leq .05$) different from one another. Approximately 13% of the variance in spring WRAT scores could be attributed to sponsor "effects".

b. No significant overall FT/NFT main effect detected.

c. No significant sponsor by cohort interaction.

2. Analysis B.

   a. No significant cohort differences.

   b. Some sponsor-by-cohort interaction was detected ($p \leq .10$). This interaction was responsible for 2.5% of the variance in spring WRAT scores.

   c. Some three-way interaction (sponsor-by-FT/NFT-by-cohort) ($p \leq .25$).

X. FT AND NFT, COHORT III PARENT CHARACTERISTICS (Vol. IB, pp. I-1 to I-80)

A. Research Questions:

1. Is there any variation by either FT/NFT overall or FT/NFT within sponsors for the following sets of variables: demographic characteristics of families, parental attitudes, and reported behavior of parents?

2. If potential mediating variables are controlled, do attitudes and reported behaviors of parents vary by FT/NFT or by FT/NFT within sponsors? (p. I-1)

B. Sample: The sample was constructed using criteria that had to do strictly with the availability of data. Parents of Cohort III EK children were distributed across sponsors as follows:

Sponsor	FT	NFT	Total
2	301	307	608
3	339	272	611
5	231	140	371
7	172	122	294
8	308	118	426
9	174	142	316
10	245	110	355
11	278	151	429
12	259	169	428
14	75	61	136
TOTAL	2382	1592	3974

C. Variables:

1. Independent:

   a. Being the parent of a FT or a NFT child;

   b. Being the parent of a child enrolled in one of the sponsor's FT or NFT classes.

2. Dependent:

   a. Income;

   b. Mother's education;

   c. Residential mobility;

   d. Mother's employment in the school.

These variables served as criterion variables for FT/NFT comparisons. In addition to these four measures, cluster analyses and judgemental comparisons were used to construct four more parent measures that served as criterion measures. These were:

e. Parent-school interaction A -- which includes the item concerning parent's employment at school;

f. Parent-school interaction B -- which excludes the item concerning parent's employment at school;

g. Parent-child, school-oriented behavior;

h. Parent satisfaction with child's academic success.

3. Covariates:

a. Parent's perception of school's receptivity to parents;

b. Parent locus of control;

c. Mother's education;

d. Parent employment at school;

e. Poverty index;

f. Parent satisfaction with child's academic success;

g. City size.

D. Analytical Method:

1. Plain descriptive statistics (e.g., medians and percentages) are used to describe the distribution of all eight parent measures across sponsors and FT/NFT groups.

2. ANOVA and ANCOVA analyses of parent-school interaction (a) and (b); parent-child, school-oriented behavior; and parent satisfaction with child's academic success.

E. Comments:

1. Variables: Two of the variables identified as outcome measures -- parent perception of school's receptivity and parent locus of control -- appear actually to have been used only as covariates.

The development of parent measures with cluster analysis was based on a different sample of parent interviews -- 13,600 kindergarten parents.

Coefficients of internal consistency for the dependent variables ranged from .42 to .60 . These low reliabilities need to be considered when attempting to interpret the reported results.

The validity of the dependent variables and both the reliability and validity of the covariates are not examined. The nature and quality of adjustments is difficult to determine.

2. Analytical Method: The ANOVA and ANCOVA results should not be interpreted as effects estimates. Given the fact that parent data were collected at only one point in time -- spring of the first year of their children's FT experience -- causal or outcome inferences are questionable.

   Although the report notes, "Tests for heterogeneity across FT/NFT by sponsor were made for the covariates used in the analysis," (p. I-42) the results of the tests are not given. The appropriateness of the ANCOVA remains undetermined.

F. Results:

1. Overall, NFT parents report higher incomes and a larger proportion of mothers completed high school and fewer mothers working in school. Little overall difference in residential mobility. Across sponsors there is variability in mobility as well as income, education and mothers working in school. Overall FT parents report more parent-school interaction (for both interaction A and interaction B), more parent-child school oriented behavior, and more parent satisfaction with child's academic success.

2. Summary statistics of regression analyses indicate that FT/NFT main effects, sponsor main effects, and sponsor X FT/NFT interaction effects together account for no more than 5% of variance in any one of the four clusters mentioned. Covariance adjustments seem to have had little impact on the overall FT/NFT or the sponsor FT/NFT comparisons, but the exact manner in which ANCOVA affects FT/NFT comparisons is not discussed.

XI. FT AND NFT, COHORT I PARENT CHARACTERISTICS (Vol. I B, I-1 to I-80)

This present analysis is parallel to the analysis described immediately above -- i.e., FT and NFT, Cohort III Parent Characteristics. Readers are therefore referred to the previous summary for details.

A. Research Questions: See previous summary.

B. Sample: The sample was constructed on the basis of the availability of data. Eight hundred and thirteen parents, who were interviewed in the spring of 1972 when their children were in the third grade, were selected. They were distributed across sponsors as follows (01 refers to self-sponsored programs):

		Parents (n)		
Sponsor	Sites (n)	FT	NFT	Total
01	2	57	19	76
05	1	55	12	67
06	3	116	41	157
07	2	50	35	85
09	2	75	76	151
10	1	57	34	91
11	1	75	41	116
12	1	46	24	70
TOTAL	13	531	282	813

C. Variables:

1. Independent: same set as in previous summary.

2. Dependent: same set as in previous study with two additional criterion measures:

   i. Parent knowledge of FT; and,

   j. Parent satisfaction with child's affective growth.

3. Covariates: a slightly different set was used.

   a. Parent's perception of school's receptivity to parents;

   b. Parent locus of control;

   c. Mother's education;

   d. Parent employment at school;

   e. Household income.

D. Analytical Method: See previous summary.

E. Comments:

1. Variables: Note that the four clusters in this study which carry the same names as those in the previous study (i.e., parent-school interaction A, parent-school interaction B, parent-child's school oriented behavior, and parent satisfaction with child's academic success) do not include identical sets of items. It is not entirely clear, however, exactly how they differ.

   Reliabilities (coefficients of internal consistency) on the dependent variables range from .48 to .80 . Results should be interpreted appropriately.

   Fallibility of covariates remains unexamined. Interpretation of adjusted results from the ANCOVA is made difficult.

F. Results:

1. Comparisons of FT and NFT parents: Overall, NFT parents report considerably larger incomes, slightly larger proportion of mothers completing high school, the same proportion (.42) of families at the same address for seven or or more years, and the same proportion (.17) of mothers working in school as FT parents.

   Across sponsors there is considerable variability on all demographic variables.

   Overall, FT parents report slightly more parent-school interaction (on both parent-school interaction clusters A and B). No difference between FT and NFT parents in amount of parent-child school oriented behavior (p. I-63). Report concludes "FT parents perceive that they are fairly knowledgeable about FT program...." (p. I-63). Yet Table M I-23 indicates that, for four of the seven main sponsors included in this analysis, NFT comparison parents reported more knowledge about FT than did FT parents themselves. Overall, FT and NFT parents seem equally satisfied with children's academic success. FT parents report more satisfaction with children's affective growth than NFT parents in general.

2. ANOVA and ANCOVA results: FT/NFT, Sponsor and FT/NFT X Sponsor interaction effects altogether account for only 3% of variance in the satisfaction with academic success variable, but 12% of variance in affective growth variable. For specific sponsors, FT/NFT comparisons vary considerably except for parent satisfaction with children's affective growth. For this variable, FT parents report more satis-

faction than NFT parents for all sponsors except, surprisingly, Sponsor 11. Covariance adjustments seem to have had little impact on sponsor FT/NFT comparisons, but the exact manner in which adjustments affect these comparisons is not discussed.

XII. PARENTS' ATTITUDES AND BEHAVIORS (Vol. I B, pp. I-1 to I-80)

A. Research Questions:

1. Is there a relationship between parental attitudes and reported behaviors, and: family income, parents' perception of school's attitudes towards parents, reported parental interactions with their children?

2. Do any such relationships vary by FT/NFT within sponsor? (see p. I-2)

B. Sample: The same Cohort I EK parent sample as was used for the FT and NFT, Cohort I Parent Characteristics Analysis.

C. Variables:

1. Independent:

   a. Parent-school interaction A;

   b. Parent-school interaction B;

   c. Income;

   d. Parent-child, school-oriented behaviors;

   e. Parent Perceptions of school's receptivity to parents.

2. Dependent:

   a. Parent-school interaction A;

   b. Parent-school interaction B;

   c. Parent satisfaction with child's academic success;

   d. Parent satisfaction with child's affective growth.

3. Covariates:

   a. FT/NFT;

   b. Eight sponsors;

   c. All two- and three-way interactions (with the various outcome measures).

D. Analytical Method: Simple regressions of each dependent variable on each independent were performed first. The multiple regressions including covariates were undertaken.

E. Comments: Most of the specific remarks made concerning the previous two parent analyses are applicable here. Great caution should be exercised in drawing inferences from these results.

F. Results: Parent-child school oriented behavior has positive (statistically significant at .005 level) regression relations with parent-school interaction (A & B) and with parent satisfaction with child's academic success and affective growth. SS positive relations remain when covariates are introduced. Parent perception of school's receptivity to parents has SS positive relationship with parent satisfaction with child's affective growth both before and after covariance adjustment. Income level has SS significant positive relationships with both parent-school interaction variables both before and after covariance adjustment. Income also has SS negative relationship with parent satisfaction with child's affective growth but this relationship disappears when covariates are introduced. Parent-school interaction A has SS relationship with parent satisfaction with child's academic success, but this relationship decreases with covariance adjustment.

XIII. TEACHER CHARACTERISTICS (Vol. I B, pp. II-1 to II-77)

A. Research Question: Do the personal and professional characteristics of teachers vary by FT/NFT, by FT/NFT within sponsor, by grade level, by community location, or by a combination of these?

B. Sample: The sample was composed of over 1,600 teachers (K, 1, 2, 3, and non-graded) from 14 sponsors, who responded to the Teacher Questionnaire mailed by SRI in Spring 1972. "Of those surveyed, 1,122 teachers (631 FT and 491 NFT) met the following criteria: (1) they taught in graded classrooms; (2) they were associated with one of the ten sponsors examined in the school studies at the kindergarten level; and (3) they had complete data on all items pertinent to the FT/NFT analyses. Data for these teachers were used to answer all questions pertaining to FT/NFT differences except where age was the dependent variable. In this instance, data from 1,084 teachers who reported their age were analyzed.

C. Variables;

1. Independent:

   a. Being a teacher of a FT or a NFT class;

   b. Affiliation with one of the sponsors;

   c. Grade level of class taught;

   d. Community location.

2. Dependent:

   a. Teacher age;

   b. Years of total teaching experience;

   c. Years of teaching experience in current school;

   d. Years of experience in sponsor's approach;

   e. Teacher's education;

   f. Teacher's salary;

   g. Teacher's ethnicity (white or nonwhite);

   h. Teacher's place of residence (in neighborhood similar or different from that of pupils).

D. Analytical Method: Cross tabulations.

E. Comments: Although the text discussing the results of this study generalize the findings of the analysis to FT and NFT teachers in general, the study provides no information on possible non-response bias. Also, no information is given on the reliability and validity of the information derived from the teacher questionnaire which is used in this analysis. Note also that this analysis examines data from teacher questionnaires administered at only a single point in time. Therefore, patterns of differences revealed in this analysis cannot with any degree of confidence be interpreted as sponsor or FT effects.

F. Results: "FT teachers are on the whole younger than NFT teachers, somewhat less experienced in general, and slightly newer to their present school assignments....on the average they are in their mid-thirties, have over nine years of total teaching experience, and over five years experience in their present schools. The FT/NFT contrasts are generally consistent across Sponsors and grade levels with the exception of Sponsor 12 (Pittsburgh), which appears to have FT teachers who are older and more experienced than their NFT counterparts, especially at the higher grade levels.

"Although there is some variation by Sponsor, on the average FT kindergarten teachers have been with their Sponsor's approach somewhat longer than teachers in the higher grades." (p. II-27)

"...age and experience are relatively independent of the size of the community or the region in which teachers are located." (p. II-32)

"...there is little difference between FT and NFT teachers on education attainment. Overall, 71.8% of the FT teachers and 68.8% of the NFT teachers have earned credits or degrees beyond the Bachelor's level...In general, the kindergarten teachers have attained more advanced credits and degrees than the teachers in higher grades. Moreover, at the kindergarten level, the NFT teachers have attained a higher level of education than the FT teachers; at the higher grade levels, the FT teachers have attained a higher level of education than the NFT teachers ...the relative educational attainment of FT/NFT teachers does vary from Sponsor to Sponsor." (p. II-32) "The higher education of the kindergarten teachers may thus be a function of their primarily non-southern, urban location." (p. II-35)

"...NFT teachers earn slightly higher salaries than the FT teachers, possibly reflecting the difference in experience. The kindergarten teachers also earn higher salaries than the teachers in other grade levels....these grade level differences are apt to reflect a geographic difference." (p. II-35)

"...there are more minority teachers in FT than NFT classes... also slightly more minority teachers in kindergarten, once again reflecting the non-southern, urban location of the kindergarten group...there is variation by FT/NFT and grade level within Sponsor." (p. II-38)

"...a slightly higher percentage of FT than NFT teachers report living in the same neighborhood or in a neighborhood similar to the one in which their pupils live...this FT/NFT contrast is generally consistent for all Sponsors across grades."

"...(See Table MII-11 for the correlations.) The larger the city, the more apt teachers are to report living in a neighborhood similar to that of their pupils." (p. II-38)

XIV. TEACHER TRAINING OF FT TEACHERS (Vol. I B, pp. II-1 to II-77)

A. Research Questions: Do the amount and type of training reported by FT teachers vary by sponsor, by grade level, by community location, or by a combination of these?

B. Sample: From the 1,122 teachers who formed the sample for the teacher characteristics analyses, "a subset of 507 teachers also had complete data on all FT relevant variables, such as training and attitudes toward FT. Data from this subgroup were used in those analyses pertaining only to FT teachers." (p. II-11)

C. Variables:

1. Independent:

   a. Sponsor affiliation;

   b. Grade level.

2. Dependent: Variables were developed from the approximately 280 items on the FT teacher questionnaire by cluster analyzing the teacher questionnaire data. Of the nine cluster variables developed, three had to do with teacher training. These three cluster variables served as the dependent variables for this analysis:

   a. Focus of teacher training: structure;

   b. Focus of teacher training: child-centeredness;

   c. Focus of teacher training: working with parents and aides.

3. Covariates:

   a. City size (except for training focused on child-centeredness outcome);

   b. South/non-south;

   c. West/non-west;

   d. Years of total teaching experience;

   e. Years of experience in sponsor's approach;

   f. Teacher education;

   g. Teacher ethnicity.

D. Analytical Method: Two-by-nine crossed factorial, ANOVA and ANCOVA models were used.

E. Comments: Also see comments in the "Teacher Characteristics" study in summary XIII, above.

Variables: The coefficients of internal consistency for the dependent variables were .71, .90, and .92, respectively. The latter two certainly measure up to conventional standards. The validity of all the variables remains undiscussed and undetermined as does the reliability of the covariates.

F. Results: "...FT kindergarten teachers in sponsors 7 and 8 report receiving a great deal more training in structure than the average kindergarten teacher. To a lesser extent, the teachers associated with Sponsors 12 and 14 also report receiving greater than average amounts of training in structured learning activities. On the other hand, the teachers in Sponsors 5, 10, and 11 report receiving far less training in this area than average and teachers in Sponsors 2 and 9 somewhat less." (p. II-40)

"...Kindergarten teachers in Sponsors 7 and 10 report receiving less training in child-centeredness,...kindergarten teachers in Sponsor 11 (EDC) report receiving far more of this type of training, and those in Sponsors 3 and 12 somewhat more, than the average FT kindergarten teacher.

"...The teachers in Sponsors 2, 5, and 7 report receiving less training in developing...skills [in working with parents and aides] than the average FT teacher; the teachers in Sponsors 8, 10, and 12 receiving more training; and the teachers in other Sponsors report receiving an average amount of training in this area." (p. II-44)

"Wherever possible, we have adjusted both geographic and teacher background characteristics in the findings reported. However, in the case of training in child-centeredness, community size was not taken into account because it is heterogeneous across Sponsors. That is, the relationship of community size and the amount of training teachers report having received in this area differ from Sponsor to Sponsor." (p. II-46)

"...In general, the FT teachers in very large communities report receiving less training in child-centeredness activities than those in medium-sized communities. Within some Sponsors, the FT teachers in very small communities also report receiving less training than others." (p. II-47)

These results, reported in the text of the report are accompanied by graphs of estimated relative sponsor "effects." Results of multiple regression analyses are provided in Appendix Tables AM II 4A and 4B. Although nowhere explicitly stated, covariate adjustments seem to have made very little difference in findings.

XV. TEACHER ATTITUDES AND BEHAVIORS (Vol. I B, pp. II-1 to II-77)

A. Research Questions: Do teachers' values, attitudes, and reported behaviors vary by FT/NFT, by FT/NFT within sponsor, by grade level, by community location or by a combination of these?

B. Sample: The same sample was used in these analyses as was used for the Teacher Characteristics analysis summarized above (N = 1122).

C. Variables:

1. Independent:

   a. Being the teacher of a FT or a NFT classroom;

   b. Affiliation with one of the sponsors;

   c. Grade level.

2. Dependent: The following cluster variables were developed through factor analysis of item information in the teacher questionnaires and then used as dependent measures in this analysis (see study summary XIV above):

   a. Teacher values: parent-community orientation;

   b. Teacher values: social skills development;

   c. Teacher values: structured/academic vs. child-centered orientation;

   d. Teacher attitudes towards meeting with parents;

   e. Frequency of teacher visits to pupil homes;

   f. Teacher expectations for pupils' academic progress.

3. Covariates:

   a. City size except for frequency of visits to pupils' homes;

   b. South/non-south (except for frequency of visits to pupils' homes);

   c. West/non-west;

   d. Years of total teaching experience;

   e. Teacher education;

   f. Teacher ethnicity.

D. Analytical Method: Crossed ANOVA and ANCOVA models were used.

E. Comments: Variables -- The coefficients of internal consistency for four of the six dependent variables were:

Teacher attitudes toward meeting with parents	(.73)
Teacher values:	
Parent-community orientation	(.54)
Social skills development	(.50)
Structure/academic vs. child-centered orientation	(.72)

The reliability of the other two dependent variables as well as the reliability of the covariates remains undetermined. The validity of none of the measures is discussed. Also see comments pertaining to Teacher Characteristics, study summary XIII above.

F. Results:

1. "When asked to choose from among competing values, FT kindergarten teachers as a group do not differ significantly from NFT kindergarten teachers in their emphasis on structured activities aimed at teaching basic skills or on the traditional early childhood goals of developing children's social skills, nor do the two groups differ significantly in the value they place on parent involvement and developing ethnic pride as primary teaching goals." (p. II-47) "However, FT kindergarten teachers do have a much more positive attitude towards meeting with parents for a variety of school-related purposes, and they make more visits to children's homes than do NFT kindergarten teachers. Finally, there is no significant difference between the FT and NFT groups in their expectations of their children's academic progress."

2. "There are small but significant differences in values, attitudes, and reported behaviors between the kindergarten teachers and the teachers at higher grade levels. The teachers at higher grade levels are more apt to value a structured academic approach over a child-centered approach to education. This is not surprising given the emphasis on reading and basic skills in these grades. On the other hand, the kindergarten teachers are slightly more apt to feel that involving parents in their children's education is a primary goal of their program, to place importance on meeting with parents, and to make frequent visits to children's homes...Finally, there is no significant grade level difference in teachers' orientations toward social skills or in their expectations for their children's

academic progress." (p. II-50)

3. "...Two of these [four] variables have significant Sponsor by FT/NFT interactions: (1) teacher values: structured/academic versus child-centered orientation and (2) frequency of teacher visits of children's homes." (p. II-50)

4. "...when asked to choose from among competing goals, the FT kindergarten teachers in Sponsors 2, 8, and 12 are somewhat more apt to place greater value on involving parents and developing ethnic pride than the NFT teachers, whereas FT kindergarten teachers in Sponsors 5 and 7 are less apt to value these goals than their NFT counterparts..."

5. "As for social skills,...FT kindergarten teachers in Sponsors 7, 9, and 11 place less value on the development of social skills than do NFT teachers, while in Sponsors 12 and 14 the reverse is true.

6. "...The FT kindergarten teachers in Sponsors 7, 8, 10, 12, and 14 value basic skills and the use of structured learning activities more than do their NFT kindergarten counterparts. On the other hand, the FT kindergarten teachers in Sponsors 3 and 9, and to some extent 11, are more child-centered than their NFT counterparts." (p. II-52)

7. "It is important to point out, once again, that these FT/NFT within sponsor contrasts are only relative...For example, the raw means on the structured/academic versus child-centered variable...show that Sponsors 2 and 5 have very child-centered NFT kindergarten teachers relative to the average NFT group at this grade level...In either case, this orientation obscures the highly child-centered values of the FT groups. It is suggested that the reader examine the descriptive statistics for all variables to more fully understand the nature of these FT/NFT within Sponsor contrasts." (p. II-56)

8. Many values profiles "bear close resemblance to the training profiles...despite the fact that neither the variables nor the analytic groups are the same" in these two studies. (p. II-56)

9. "...teachers' orientations toward structured/academic versus child-centered instruction also vary by FT/NFT depending on grade level within Sponsors...in most cases, the higher a teacher's grade level, the more structured her orientation is likely to be. However, this pattern is reversed in Sponsor 9, whose NFT teachers value structure less in higher grades than in kindergarten." (p. II-58)

10. Overall FT teachers report higher "frequency of...visits to children's homes, [but]...there is wide variation among Sponsors. FT kindergarten teachers in Sponsors 3, 9, 12,

and 14 make far more home visits than do NFT kindergarten teachers, in Sponsor 7 they make somewhat fewer, and in Sponsors 5, 8, 10, and 11 they make about the same number of home visits as their NFT counterparts. It should be pointed out, however, that Sponsor 10's NFT group makes far more home visits ($\bar{x}$ = 5.06) than the average NFT group ($\bar{\bar{x}}$ = 2.76).

11. "For reasons of heterogeneity, it was impossible to adjust this variable for community size or Southern location, both of which are related to the frequency with which *these* teachers make home visits." (p. II-58)

12. Results of covariance adjustments are not explicitly recounted but charts in the text and tables indicate that the introduction of covariates made little difference in results.

XVI. TEACHER SATISFACTION WITH AND FAITHFULNESS TO SPONSOR'S MODEL (Vol. I B, pp. II-1 to II-77)

A. Research Questions: Does either teacher satisfaction with or perceived faithfulness to the sponsor's approach vary by sponsor, by grade level, by community location, or by a combination of these? (p. II-2)

B. Sample: These analyses utilized the limited subset of teachers (N = 507) who had complete data on all FT relevant variables.

C. Variables:

1. Independent:
   - a. Sponsor;
   - b. Grade (K or other grade teacher).
2. Dependent: Two cluster variables developed from the teacher questionnaire (see study summary XIV above):
   - a. Teacher satisfaction with sponsor approach;
   - b. Perceived faithfulness to sponsor approach.
3. Covariates:
   - a. City size (except for teacher satisfaction variables);
   - b. South/Non-south;
   - c. West/Non-west;
   - d. Years total teaching experience;
   - e. Years experience in sponsor's approach;
   - f. Teacher's education;
   - g. Teacher ethnicity.

D. Analytical Method: Crossed ANOVA and ANCOVA models were used.

E. Comments:

1. The coefficients of internal consistency for the two dependent variables were: teacher satisfaction (.76) and perceived faithfulness (.80).

   The relability of covariates as well as the validity of all measures remain undiscussed and undetermined.

2. Perceived faithfulness of sponsor approach should be understood for what it is -- i.e., a measure of how closely to the model FT teachers feel they adhere to their sponsor's model. It should not be taken as an objective measure of classroom activity.

3. See comments in study summary XIII above.

F. Results: "With a possible range of 4 to 17 points on the satisfaction variable, FT teachers overall have a mean raw score of 15.00 and a standard deviation of 2.99 . Thus, to the teachers who work in the programs, FT appears to be extremely satisfying personally and worthy of recommendation to others. The faithfulness variable has a possible range of 9 to 33 points. With a mean raw score of 26.33 and a standard deviation of 5.13, FT teachers also feel that their classrooms reflect the Sponsor's ideal.

"...kindergarten teachers appear to be slightly more satisfied than teachers at higher grade levels, both before and after adjusting for regional and teacher background variables... there is much variation in teacher satisfaction among Sponsors at the kindergarten level. While it must be emphasized again that these variations are relative only, and that overall teachers are highly positive in their attitudes, it is clear that certain Sponsors engender somewhat more satisfaction than others. When region and teacher background are held constant, teachers in Sponsors 3, 11, and 12 report relatively more satisfaction, and teachers in Sponsors 7, 9, and 14 relatively less..."

"The degree to which teachers are satisfied with their Sponsor's approach is further affected by city size. In general, the bigger the city the less satisfied the teachers...The Sponsor comparisons above, however, are not adjusted for city size because relationship between city size and satisfaction varies from Sponsor to Sponsor." (pp. II-63, II-65)

"...Kindergarten teachers feel that their classroom is somewhat closer to their Sponsor's ideal classroom than do teachers at higher grade levels. Teachers' perceived faithfulness also varies among Sponsors, especially when one fails to adjust for city size, region and teacher background variables. With these adjustments, three Sponsors appear to fall outside the norm. FT kindergarten teachers in Sponsors 10 and 12 feel more faithful to their respective models than the other and teachers in Sponsor 9 somewhat less so. Clearly, site variability must once again be taken into account.

"...Teachers' perceptions of faithfulness to their Sponsor's approach vary not only by Sponsor and grade level, but also by grade level within Sponsor. Most kindergarten teachers perceive themselves as being more faithful to their Sponsor's approach than teachers at higher grade levels. Kindergarten teachers in Sponsors 3 and 8, however, feel less faithful to their respective models." (p. II-68)

In general, judging from charts on pp. II-67 and II-69 and Appendix Tables AM II-4A and AM II-4B, covariance adjustments seem to have resulted in findings no different from those of unadjusted analyses.

XVII. INTERRELATIONS OF TEACHER CHARACTERISTICS (Vol. I B, pp. II-1 to II-77)

A. Research Questions: Are teachers' personal and professional characteristics related to their reports of training, values, attitudes, and reported behaviors? Do these relationships vary by FT/NFT, by FT/NFT within sponsor or by grade level? (pp. II-2)

B. Sample: The sample of 1122 teachers used for the Teacher Characteristics analyses was used in this analysis as well.

C. Variables: All 18 teacher variables mentioned thus far served as both predictor and criterion variables. These were:

1. City size;
2. South;
3. West;
4. Total experience;
5. Education;
6. Ethnicity;
7. Experience with approach;

[Note: variables 8-16 in this list are cluster variables developed from cluster analysis of teacher questionnaire data. Numbers in parentheses indicate coefficients of internal consistency of these cluster variables. See p. II-22.]

8. Satisfaction (.76);
9. Perceived faithfulness (.80);
10. Training: structure (.71);
11. Training: child-centeredness (.90);
12. Training: working with parents and aides (.92);
13. Attitude toward meeting with parents (.73);
14. Values: parent orientation (.54);
15. Values: social skills (.50);
16. Values: structure/academic vs. child-centeredness (.72);
17. Frequency of home visits;
18. Teacher expectations of academic progress.

D. Analytical Method: Zero-order and partial correlations were undertaken.

E. Comments: See Teacher Characteristics study summary XIII above.

F. Results: The results of the zero-order correlations are summarized in Table MII-28 of the next page and are commented on below.

Although the results for some partial correlations are mentioned in the text of the report, the report provides no overall description of the results from the partial correlation analysis.

With respect to Table MII-28, the text of the Abt report notes: "There were no significant differences in the relationships among the two sets of variables by FT/NFT, by FT/NFT within Sponsor, or by grade level. Therefore, all teachers have been treated as a single group, and simple and partial correlations have been used to explore overall relationships.

"What is perhaps most striking about Table MII-28 is the relatively small size of the relationships observed between teachers' personal and professional characteristics and their values, attitudes and reported behaviors." (p. II-72)

Table MII-28

Interocorrelation Matrix of Community Related Variables, Teacher Background Characteristics, Teacher Training and Teapher Values, Attitudes, and Behaviors [a]

Variable Name	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1. City Size		-.34	-.01	.01	.24	-.26	.07	-.14	-.02	-.20	-.15	-.07	.09	.15	.00	.01	-.11	.04
2. South			-.26	.02	-.19	-.06	-.12	.02	.05	.03	.06	.14	-.04	-.03	-.05	-.03	.11	-.14
3. West				.02	.12	.08	-.03	.05	.08	.00	.06	.00	-.03	.02	.00	.02	.08	.03
4. Total Experience					.16	-.14	.38	.12	.14	.16	.23	.25	-.01	.04	.01	.03	-.04	-.02
5. Education						-.14	.13	-.07	.05	-.07	-.08	-.06	.06	.04	.00	.02	-.02	.05
6. Ethnicity							-.03	.00	-.07	-.15	-.13	-.15	-.08	-.23	.17	-.13	.00	-.14
7. Experience with Approach*								.12	.15	.05	.14	.10	-.16	.02	-.03	-.07	-.01	.04
8. Satisfaction*									.27	.22	.44	.24	-.16	.06	-.03	-.12	.16	.11
9. Perceived Faithfulness*										.22	.22	.24	-.11	.04	-.11	.00	.07	.04
10. Training: Structure*											.45	.39	-.09	.06	-.14	.24	-.02	.05
11. Training: Child Centeredness*												.57	-.19	.09	-.05	-.11	.08	.10
12. Training: Working with Parents and Aides*													-.21	.21	-.02	-.07	.10	.10
13. Attitude Toward Meeting with Parents														.27	-.02	-.12	.11	.03
14. Values: Parent Orientation															-.05	-.06	.10	.10
15. Values: Social Skills																-.14	.03	.01
16. Values: Structure/Academic vs. Child Centeredness																	-.06	.00
17. Frequency of Home Visits																		-.01
18. Teacher Expectations of Academic Progress																		

[a] N = 1122 FT and NFT teachers except where asterisked.
For these variables N = 507 FT teachers only.

"...there is a small positive correlation between length of total teaching experience and FT teachers' perceived faithfulness to, and satisfaction with, their Sponsor's approach. Since total teaching experience was found to be positively related to length of teaching experience in a Sponsor's approach, this latter variable was partialled out of the correlation. The partial correlations are .09 and .08 respectively..." (p. II-68)

"There is also a small positive correlation between length of total teaching experience and the amount of training teachers report receiving in three different content areas: structure, child-centeredness, and working with parents and aides. It is clear that the longer a teacher is associated with FT, the more training she is likely to have. If we hold this variable constant (via partial correlation), the correlations between experience and amount of training reported remain small, positive and significant. (The partial correlations are .15, .20, and .23, respectively.)...

"Turning to ethnicity, there is also a small correlation between the race of these teachers and the amount of training they report having received. White teachers report having received less training than minority teachers in all three areas. Even after partialling out community size and total teaching experience (both of which are related to ethnicity and the amount of training reported), this relationship remains. (That is, -.19, -.15, and -.15, respectively)...

"Teacher ethnicity is also associated with teacher values and expectations. Compared to white teachers, minority teachers are somewhat more apt both to have high expectations of their children's academic progress and to value working with parents and developing ethnic pride in their pupils...They also tend to place less emphasis on social skills development than do White teachers. These relationships are remarkably consistent across FT/NFT, Sponsors, and grade levels." (p. II-71)

XVIII. IMPLEMENTATION (Vol. I B, III-1 to III-35)

This summary departs from the format used in summarizing the other analyses in Abt's 1974 report because Abt's investigation of the implementation issue was very different than other studies summarized above. We provide simply a descriptive summary on efforts to determine "what type of inputs can reasonably be expected to contribute to the construction of scales of sponsor implementation?" (p. III-4)

The inquiry focused on five sponsor-sites in Philadelphia and four in New York City:

Philadelphia: Bank Street Model; Kansas Behavior Analysis Model; Florida Parent Educator; SEDL Model.

New York: Bank Street Model; University of Oregon Engelmann/Becker Model; High/Scope Cognitively Oriented Curriculum; University of Kansas Behavior Analysis Model.

The methods of investigation for this study were interviewing and review of written materials. However, "Sponsors varied in procedures they used in maintaining liaison (some used discipline-based consultants, some grade-level consultants, and others locally-based professionals versed in the model). ...Thus the content of field reports varies considerably across Sponsors. To avoid these problems...an alternative was adopted. The FT Director or the person employed by the LEA to implement the delivery of a specific FT program, or both if available, were interviewed." (p. III-4)

Areas of questioning concerned:

--Program or sponsor selection and assignment within district (2 questions);

--Relationship between sponsor and LEA (5 questions);

--Teachers: assignment, training, characteristics (6 questions);

--Relations and roles of school principals (3 questions);

--Local community control and role of parents (2 questions);

--LEA problems or crises (6 questions);

"The technique used to obtain the study data was a semi-structured interview...with either the FT Director or the Program Coordinator... Overviews of the FT Program in the cities being studied were obtained from the Directors of the LEA's Early Childhood Program and the FT Director for the city. Where the interviewee permitted, the conversations were taped and later transcribed verbatim. The interviewee reviewed the texts, and edited out any personal comments or asides. Only one of the reports presented here was taken from notes because the Program Coordinator did not wish to be taped. The initial conversations in Philadelphia were carried out by two interviewers to establish a consistent format. The remaining interviews were recorded by one interviewer." (p. III-6)

General results for each of the areas investigated are reported as follows:

> Philadelphia: "Immediately prior to FT, Philadelphia Schools had a five-year Education Improvement Project (EIP), which brought many new ideas and materials into the ghetto schools. However, their 'improvements' were primarily within the traditional subject-matter, teacher-controlled, classroom context. Kindergartens, on the other hand, were considered 'free of subject matter requirements,' according to the 1960 Philadelphia curriculum guide, and the major focus was on the needs of the child.
>
> "In September 1968, a formal resolution was passed by the Philadelphia Board of Education approving the FT program but constraining Sponsor selection to five specifically named national Sponsors and two local or 'self-sponsored' models. The presentation of these FT options was made to the Superintendent of Education, the eight District Superintendents, other Central Office personnel, and some of the principals. Two of the Philadelphia school districts with primarily white, middle class populations were ineligible for FT. The Superintendent wanted each of the remaining six districts to select a different Sponsor, with three schools in a district having the same program. However, at that time there was a great deal of pressure for decentralization, and the District Superintendents insisted on making their own decisions. Several districts elected to have more than one Sponsor, and not all districts had the same number of eligible schools.
>
> "The guidelines for school eligibility were basically economic. At least 50% of the children had to have been in Philadelphia Head Start ("Get Set") Centers which fed into the elementary schools, and sufficient classroom space had to be available to permit a top limit of 25 children in kindergarten and 30 children in first grade. In addition, the kindergarten programs had to provide a full day experience. Within these ground rules the District Superintendents selected schools for inclusion in FT. Furthermore, although the school staff and parents had some opportunity to express their preferences, it was the District Superintendent who made the final decision.
>
> "At the outset the distribution of FT schools was as follows:

District 1: 3 Philadelphia Process Model;
District 2: 1 Kansas, 1 Bank Street, 1 SEDL;
District 3: 1 EDC, 1 Florida;
District 4: 2 Kansas;
District 5: 2 Bank Street, 2 SEDL, 1 Parent Model; and,
District 6: 3 EDC.

"Toward the end of the first year (1968-69), the FT administrators came to Philadelphia and expressed deep concern about the manner in which the schools had been allocated to the various Sponsors. As a result of a series of meetings with the Central Office staff and the District Superintendents, it was decided to decentralize the administration of the FT programs. A few changes in allocation of schools to Sponsors also occurred at this time. A separate budget was set up for each of the 18 schools and one for the Central Office.

"One of the most crucial impacts of FT in Philadelphia is the emphasis on a totally new orientation to teaching. In the past Philadelphia has been subject-oriented, but now the schools are faced with a variety of program models based on different learning theories, focusing on the entire process of teaching rather than on curriculum content.

"Institutional change is a basic goal of the FT program in Philadelphia. Along with the notion of decentralization, the Superintendent is attempting to develop procedures to carry on the most fruitful innovations after the present federal funding is terminated. Primary agents in the change process are the FT Program Coordinators. They were selected from among the personnel of the EIP program when it was disbanded; each Sponsor was assigned his own Project Coordinator. Presumably some attempt to match people with programs was made, with appropriate retraining in the basic tenets of the model, so that now Coordinator is an advocate for that Sponsor.

"During the 1972-73 school year, the Philadelphia Schools suffered a long and bitter teacher strike. The extent of involvement in this conflict varied considerably across the FT schools. The hostilities engendered among those who participated and those who crossed picket lines endured long after the strike was over. In general, most of the FT schools remained open, with parents and other community personnel replacing the absent teachers. This situation was particularly devastating to programs whose major focus was working with parents." (pp. III-8 - III-10)

New York: "New York City was one of the earliest recipients of FT fundings, and most of the present FT schools participated in the 1967-68 school year. However, there were no Sponsors and the FT program consisted primarily of some type of open, child development approach. In the interview with the Central staff it became apparent that this is the primary orientation of the New York City FT office. All the personnel, including the Early Childhood Program Director, the FT Director, and her assistants, have a strong bias in favor of this type of educational approach.

"The procedure for Sponsor selection in New York City was very different from that used in Philadelphia. From all reports it was very similar to an auction sale.

"The introduction of FT in New York City came close on the heels of the big strike of 1968, when the decentralization movement was at its peak. The pressure for community control of educational decisions meant that a great deal of attention was paid to the local preferences. The New York City Early Chilhood Coordinator, the District Superintendents, and the Deputy Superintendent decided which schools would receive FT funding. A basic criterion was the availability of a large number of children (150 or more) with preschool experience. These were the children who went into the FT kindergarten classes. New kindergarten entrants went into the traditional kindergarten. However, these classes were also receiving federal assistance from a program called "Strengthening Early Childhood" which was supported by Title I funds. This program provided one aide for every classroom; the Follow Through funds provided a second aide for the classes in that program. Another difference between the regular and Follow Through kindergartens was that FT provided an additional one-and-one-half hour of instruction per day.

"In the Spring of 1968, a group representing the designated schools, parents representing children scheduled to enter kindergarten in those schools, the city-wide PAC organization, and LEA administrators met at the Board of Education, where they were told about FT and the Planned Variation concept. They were then introduced to the eight Sponsors who were present. The models included the Bank Street College Approach, University of Kansas Behavior Analysis Model, High/Scope Cognitively Oriented Curriculum, Florida Parent Educator Model, Hampton Institute Non-Graded Model, University of Oregon Engelmann Becker Model, Southern University Home-School Partnership, and New York University Interdependent Leaning Model. It may be that other Sponsors such as Arizona, Pittsburgh, Far West, etc., were not at this meeting because they were already over-committed to other sites. Whatever the reasons for inclusion or exclusion, the reports from the FT

> Director's Office and from the various Program Coordinators concur in identifying the group of Sponsors 'showing their wares' at the meeting.
>
> "After all the Sponsors had presented their programs materials, and educational philosophy, the representatives from the different schools were told to go back to their communities and discuss the various options. Then they were to come back to another meeting and final assignments would be made.
>
> "Just as in the business world, the law of supply and demand was operative. The constraint was the amount of money available; that is, each Sponsor was limited to one school. Thus once a Sponsor was chosen by one school it could not be selected by any other." (pp. III-23 - III-24)
>
> In addition to these overall findings, the study reports on each of the nine sites investigated. Though only briefly recounted, these findings make it clear that the implementation of the various curriculum models was strongly affected by the characteristics of local institutions and personnel.
>
> The study concludes: "The information provided by this implementation study is obviously anecdotal and requires replication and validation to have any usefulness for inclusion in the analyses of outcomes. However, the events reported have raised some intriguing questions. Will any of the idiosyncratic and 'unplanned' variations which have been documented here actually have sufficient impact to affect differentially the outcomes of the 'planned' variations? ...No attempt should be made to relate specific anomalies as they turn up in the data analyses on a one-to-one basis with the material revealed in the interviews. However, the 'real life stories' do offer some clues for deeper insights into the human variables in program delivery than can be obtained from the most sophisticated quantitative analyses." (p. III-35)

* * * * *

We would offer only two additional comments. First, although providing much interesting information concerning FT implementation, little specific information was obtained that could be directly applied to the question raised initially. Little of the preceding casts any light on the procedures to be followed in constructing scales to measure sponsor implementation. Second, this information indicates how the implementation of FT programs can be highly influenced by local events, many of which lie beyond the control of the evaluator, the curriculum model sponsor, the educational administrator and the teacher.

XIX. A POTPOURRI OF ODDS AND ENDS

Scattered among the 18 analyses summarized above were a handful of tangential inquiries which we summarize here.

A. Socioeconomic Status and Pupil Achievement (Vol I A, pp. VII-45 to VII-48): The primary question investigated was whether the correlations between SES measures and academic achievement would decrease over the period fall to spring of the kindergarten year.

The full sample of 288 schools used for the One-Year Kindergarten Outcomes analyses was used here.

Three SES measures were used -- adjusted income, percentage of tested children in the school who are Black, and mother's education. All measures are school measures. Three zero-order correlations were calculated for the fall and another three for the spring. Differences between respective correlations were tested. The results are reported in Table VII-8.

The inference drawn from these results is that "...the data can be taken as supporting evidence that FT, in the aggregate, is providing positive effects." (p. VII-48)

Such an inference is based on the tacit expectation that FT experience will in the course of a single year somehow break the characteristic association between academic achievement and SES. If FT is truly compensatory, then children with low SES background will advance sufficiently to do as well as children from high SES backgrounds.

The logic is relatively uncontestable, but it is difficult to see how the methodology has much to do with the logic. Changes in zero-order correlations of aggregated data are likely to reflect much more than the associations between the particular variables of interest. Consequently, comparison of pairs of correlations are very hard to interpret. In the absence of evidence concerning distributions and change scores at both time periods and in the absence of information about the validity and reliability of the various measurements, it's difficult to determine what can be made of the reported results.

Table VII - 8

ZERO ORDER CORRELATIONS BETWEEN SES INDICATORS AND FALL AND SPRING WRAT

FT Schools N = 156	Fall WRAT		Spring WRAT		P Differences
	r	r^2	r	r^2	
Adjusted income level	.4675	.2185	.2931	.0859	<.025
% Minority	-.2985	.0891	-.1132*	.0128	<.025
Mother's education	.4748	.2254	.3309	.1095	<.025

NFT Schools N = 132	Fall WRAT		Spring WRAT		P Differences
	r	r^2	r	r^2	
Adjusted income level	.7615	.5799	.6034	.3641	<.001
% Minority	-.6216	.3864	-.5195	.2698	<.025
Mother's education	.7213	.5202	.5305	.2811	<.001

* This correlation is significant at the .08 level. All other correlations are significant at the .01 level.

B. Pre-test Delay (Vol. I A, pp. VII-27 to VII-29): Since some children were tested later than others in the fall of 1971, there was some concern that fall WRAT scores might reflect sponsor effects when tests were delayed. The sample used to investigate this problem was the sample of 251 Cohort III schools (i.e., excluding big city schools) used for the One-Year Kindergarten Outcomes analyses.

Zero-order correlations were calculated between Fall WRAT scores and the number of days from the beginning of school until the administration of the WRAT exam for the FT and NFT children of each sponsor. The results are given in Table VII-5.

The results are difficult to interpret for a host of reasons. We mention only a few.

1. Comparison of FT with NFT correlations in order to detect differences in correlations, and infer from them the influence of test delay, assumes that FT and NFT groups are comparable initially. In light of the administrative and political nature of the selection of FT and NFT sites, this is not a very likely assumption.

2. Such a comparison also assumes that the only significant different factor operating to influence test scores during the test delay is the sponsor treatment. In light of the lack of demonstrated initial comparability and in light of of uncontrolled historical events which could differentially influence either or both groups, this assumption cannot easily be defended.

3. Even if FT and NFT groups within sponsors were initially comparable and extraneous events were controlled, differences between correlation coefficients would be difficult to interpret.

4. The report does not indicate the actual time factor involved. Was the delay a day, a week, a month, or longer? In the absence of this information it is dangerous to place much confidence in the inference that positive correlations indicate model effects.

TABLE VII - 5

Zero-Order Correlations Of Pretest Delay with Fall WRAT Scores By Sponsor by FT/NFT

	FT		NFT	
Sponsor	Correlation	p*	Correlation	p*
2	-.4868	.01	-.6388	.00
3	-.3410	.10	-.3434	.14
5	.4735	.06	-.2317	.42
7	.6335	.04	.5994	.07
8	-.1245	.61	-.0596	.83
9	-.6206	.08	-.3527	.32
10	.2406	.37	-.4634	.21
11	.2282	.45	.2272	.48
12	-.2213	.57	.0272	.94
14	-.2127	.58	.0590	.89

[*] Probability levels are based on two tailed significance tests.

C. Time of Testing (Vol. I A, pp. VII-48 to VII-52)

Abt was interested in knowing whether spring test scores varied in relationship to the length of the period between the fall testing and the spring testing. Data concerning Cohort III EK children in the small sample used for the analysis of One-Year Kindergarten Outcomes (i.e., the smaller sample of 251 which did not include big city schools) was used in this analysis. Partial correlations between the time interval (as the predictor variable) and each of the seven outcome measures (WRAT, MAT-sounds, MAT-reading, MAT-arithmetic, Gumpgookies, Locus of Control positive, and negative) with the influence of fall WRAT scores partialled out were calculated.

The results are presented in Table VII-10 below. We concur with the report's observation that results across sponsors were varied: "...it is clear that no simple relationship exists between the several testing intervals and the scores generated during those intervals." (p. VII-52)

Table VII - 10

Partial Correlations of Spring Scores with Length of Instructional Interval Controlling for Fall WRAT NFT Schools[a]

SPRING SCORES	2 N=20	3 N=20	5 N=14	7 N=10	8 N=15
Spring WRAT	.1153	.3198	.2631	.0953	.3128
MAT Sounds	-.2257	.0496	.4394	.5301	.2649
MAT Read	-.0928	.1080	-.0364	.2351	.1904
MAT Arith.	.1429	-.1583	.6774*	.2500	.3982
GUMP-GOOKIES	.1825	-.1542	.4578	-.4549	.6929*
LOCUS OF CONTROL Positive	-.0588	.2706	-.1933	-.1067	-.3239
LOCUS OF CONTROL Negative	.1312	.2773	-.0187	.1581	-.3930

*These correlations are significant at the .10 level, using two-

[a] The sample of NFT schools for this study was 129, 3 schools less These three schools are not included here because of insufficient

SPONSOR				
9 N=10	10 N= 9	11 N=12	12 N=11	14 N= 8
-.4061	.4329	.1477	-.6257*	-.3440
-.1196	.1029	.4838	-.4007	.0502
-.0562	.0225	.0378	-.7641*	.3610
.1862	.3014	.6282*	.4723	.0114
-.5783	.8709*	.6790*	-.3429	-.4971
.1280	-.4619	-.0835	.6388*	.2658
-.1214	-.5339	-.1461	.5446	-.4558

tailed significance tests.
than the total analytic subset.
time of testing information.

1.3 Marvin G. Cline, et al. Final Report. Education As Experimentation. Evaluation of the Follow Through Planned Variation Model. Volume IIA: Two-Year Effects of Follow Through and Volume IIB: Monographs and Appendices. Cambridge, MA: Abt Associates, May 1, 1975.

Abt's 1975 report is the second in a series of Abt evaluation reports on project Follow Through. In contrast with Abt's 1974 report, it focuses exclusively on Cohort III, entering kindergarten (that is, entering kindergarten in Fall 1971) children. The latest data on child outcomes for this group of children were collected in the spring of 1973 at the end of their second grade. There is thus the potential to examine end of kindergarten effects and cumulative two-year effects. The report also attempts to identify effects attributable to the first grade experience alone.

The 1975 report also differs from the 1974 report in that it uses a different set of pupil outcome measures to estimate two-year pupil outcomes. Instead of the eight pupil outcomes (four academic, three affective, and absence) of the earlier report, only three academic achievement measures are used in the 1975 report: MAT-word analysis, MAT-mathematics, and MAT-total reading.

No parent focused inquiries are attempted in the present report -- another point of difference between this and the previous year's report.

We have summarized each of the relatively separate analyses of the report under individual titles. The titles under which the different analyses are summarized have been changed slightly from those used in the Abt report in order to make them identical to references in the report. The similarity of titles, the substance of the commentary as well as page references should leave no doubt as to which set of analyses are being summarized.

All in all, eight separate analyses are summarized:

Volume IIA

I. SAMPLE DEVELOPMENT: REPRESENTATIVENESS OF THE 1974 ANALYTIC SAMPLE . . . . p. 140
II. ONE-YEAR, KINDERGARTEN STUDY . . . . p. 149
III. TWO-YEAR EFFECTS . . . . p. 152

Volume IIB

IV. INTERIM ANALYSIS OF LONGITUDINAL FT/NFT SITE-LEVEL DIFFERENCES IN ACADEMIC ACHIEVEMENT . . . . p. 161
V. TEACHER STUDIES . . . . p. 173
VI. FT AND NFT SCHOOL CHARACTERISTICS . . . . p. 182
VII. FOLLOWING THROUGH ON HEAD START VARIATIONS: STUDY I . . . . p. 187
VIII. FOLLOWING THROUGH ON HEAD START VARIATIONS: STUDY II . . . . p. 190

I. SAMPLE DEVELOPMENT: REPRESENTATIVENESS OF THE 1974 ANALYTIC SAMPLE (Vol. II A, pp. II-1 - II-39)

Abt's 1975 report was concerned with the representativeness of the samples used for analyses. In particular, the Abt investigators sought to answer the research question: Has the application of a number of sampling restrictions led to the selection of a biased sample for effect analyses? This question was addressed by comparing different groups of students. Three comparisons were undertaken. The methods and results of the comparisons are reported shortly. However, it is necessary to say a word or two about the terms the report uses in referring to the several samples it deals with.

The section of the 1975 Abt report dealing with sample development mentions at least a dozen different samples on subsets of children whose data has been analyzed in different portions of the report.

Unfortunately, the report does not make clear the exact nature of these different subsets of children nor exactly how they relate. Confusion arises from three sources: 1) the number of different samples mentioned; 2) the use of different labels to describe the same (or at least very similar) subsets of children; 3) ambiguity over the rationale behind the comparisons employed. Nevertheless we have tried to sort out the samples for the main comparisons as best we could. The three main samples discussed are:

Initially tested CIII children (n = 17,721): children for whom initial test data is available for the point at which they entered FT, that is, kindergarten in fall 1971.

1973 Analytic Sample (n ≈ 6,000, but no single figure given): the subsample of the initially tested CIII children whose test scores as of the end of kindergarten in spring 1972. (This is apparently the group whose data was analyzed and reported in Abt's 1974 report. The "1973" in the label apparently derives from the fact that Abt received the end-of-kindergarten CIII data in 1973.)

1974 Analytic Sample (n = 2,755): the subset of initially tested children for whom data is available at the finish of first grade, that is, spring 1973. The labels ANCOVA Analytic Sample and Traced Longitudinal Sample apparently are also used to describe this sample -- or ones highly similar to it.

Other samples are described in the portions of the summary reporting the three sample comparisons grouped under Abt's "Sample "Development" rubric.

"Comparison of Traced Longitudinal Random Sample with Random Sample of Fall 1971 Tested Population" (pp. II-11--II-14)

A. Research Question: Are students for whom a complete set of outcome measures are available, representative of the entire group of FT and NFT children who were tested Fall 1971?

B. Sample: From the 17,721 initially tested Cohort III children, 6,442 entering kindergarten children were randomly selected to examine the reliability and validity of Fall 1971, Spring 1972, and Spring 1973 outcome measures. These 6,442 constituted a reference group for the creation of the two groups used for the present comparison.

"Traced Longitudinal Random Sample": From the 6,442 children selected for reliability and validity analyses, 1883 (1215 FT, 668 NFT) were identified as having a complete set of Fall 1971, Spring 1972, and Spring 1973 scores. These are referred to as the "Traced Longitudinal Random Sample."

"Random Sample": This sample appears to be composed of children among the 6,442 who did not have a complete set of Fall 1971, Spring 1972, and Spring 1973 scores.* This group of children is called the "Random Sample" and constitutes the reference group for determining whether the "Traced Longitudinal Random Sample" is representative.

---

*We say it "appears to be composed" of such children because the report is somewhat unclear in referring to this group. At one point (p. II-11) the report states that the Traced Longitudinal Random Sample is compared "with the total random sample that was tested in the Fall of 1971." We presume this refers to the 6,442 randomly drawn for reliability and validity checks. If that assumption is correct, the statement that this group constitutes a comparison group for the Traced Longitudinal Random Sample is clearly not true. The sample size of the group with which the Traced Longitudinal Random Sample is compared is a good deal smaller (4,848 in one comparison, 4,742 in another) than the "total random sample" of 6,442.

The matter is further complicated by uncertainty as to exactly how many children are in either the Traced Longitudinal Random Sample or the Random Sample. The sum of the number of children reported as being in the "Random Sample" exceeds the number of children in the sample from which both analytic samples were drawn. Figures reported in conjunction with the results of comparisons on the WRAT Scores (see Variables below) are: "Traced Longitudinal Random Sample" N = 1883; "Random Sample" N = 4848; $\Sigma > 6442$. Figures reported in conjunction with the results of comparisons on the PPVT scores (see Variables below) are: "Traced Longitudinal Random Sample" N = 1849; "Random Sample" N = 4742; $\Sigma = 6,591 > 6,442$.

C. Variables: Fall 1971 scores on two tests were used as the basis for comparing these two groups -- the Wide Range Achievement Test and the Peabody Picture Vocabulary Test.

D. Analytical Method: The mean (on both measures) of FT children of one group was compared with the mean of FT children in the other group. Similarly, the mean on both measures of NFT children in one group was compared with the mean of NFT children in the other group. Differences greater than two standard errors, and greater than one-quarter of a standard deviation were noted.

E. Comments: The comparisons of the "Traced Longitudinal Random Sample" with the "Random Sample" does address the question whether the application of sample restrictions was biasing. However, it only indirectly addressed the question of whether the application of sampling restrictions led to a biased sample in the case of the ANCOVA Analytic Sample (also apparently called the 1974 Analytic Sample). In order to draw any conclusions concerning the latter -- and definitely more important -- question, one would have to assume that the "Traced Longitudinal Random Sample" was representative of the ANCOVA Analytic Sample. This assumption is never examined. Since it is not, the comparison performed and reported can at best provide only indirect evidence that the ANCOVA Analytic Sample was or was not biased by the application of sampling restrictions.

The reference "Comparisons of Traced Longitudinal Random Sample with Random Sample of Fall 1971 Tested Population" is an obviously misleading title. The "Traced Longitudinal Random Sample" is random only in the sense that it is part of a sample randomly drawn for other purposes. Its selection is not the product of any tracing procedure. The "Random Sample" does not refer to a random sample of all, but to a subset of a random sample that was left after elements were drawn on an arbitrary basis to form the "Traced Longitudinal Random Sample."

F. Findings: On the basis of WRAT scores:

1. For seven FT sponsors this comparison revealed no "significant differences" between FT children in the "Traced Longitudinal Random Sample" and FT children in the "Random Sample" or between NFT children in the two groups.

   These sponsors were:

   Far West Labs
   U. of Arizona
   Bank Street
   U. of Oregon
   U. of Kansas
   High/Scope
   EDC

2. Traced Sample had significantly higher scores than Random Sample for both FT and NFT: U. of Pittsburgh.

3. Traced Sample greater than Random Sample for FT, but not for NFT: U. of Florida.

4. Traced Sample less than Random Sample for FT (and for NFT but not significant in the latter case): SEDL.

On the basis of PPVT scores:

1. No "significant differences" on either Traced FT with Random FT or Traced NFT with Random NFT:

Far West Labs	U. of Oregon
U. of Arizona	High/Scope
Bank Street	EDC

2. Traced Sample greater than Random Sample for both FT and NFT: U. of Pittsburgh.

3. Traced Sample greater than Random Sample for FT, but not for NFT: U. of Kansas.

4. Traced Sample greater than Random Sample for NFT, but not for FT: U. of Florida, SEDL.

In general, the report concludes: "Whatever the reasons... these findings suggest that for various Sponsor/treatment groups differential attrition does exist as we generate a traced longitudinal sample. Our first step in the sample defining process seriously complicates our ability to generalize our findings." (p. II-14)

"Comparison of 1974 Analytic Sample with [the] Traced Longitudinal Sample" (pp. II-14 to II-22)

A number of "sampling restrictions" were applied in constructing the sample used to test for overall and sponsor effects. That is, a number of conditions were set down concerning the type of student to be included in the analytic sample. Table II-1 lists most of these sampling restrictions and indicates the resultant reduction in sample size.

TABLE II-1 (from p. II-3)

SUCCESSIVE RESTRICTIONS PLACED ON ANCOVA ANALYTIC SAMPLE

Restriction	Resultant N
Fall 1971 WRAT	17721
Spring 1972 WRAT Spring 1972 MAT Spring 1973 MAT	4962
First grade status in 1972-73	4742
Remain in original treatment group	4625
Retention in same school	4467
N ≥ 5 in school	4418
[Students with no zero scores on outcome measures	4270]
Complete set of dependent measures	3665
Complete set of covariates	2790
N ≥ 5 in school	2755

A. Research Question: The present comparison attempts to determine whether the application of successive sampling restrictions resulted in biased samples as the sample was reduced from 4,418 children to the 2,755 children in the 1974 Analytic Sample.

B. Sample: From among Cohort III, Entering Kindergarten children, five "samples" were "constructed":

(1) 4,418 pupils still being considered before the application of a "no zero scores on outcome measures" restriction was applied;

(2) 4,270 pupils who remained after the exclusion of those with zero scores on outcome measures;

(3) 3,665 pupils who had a complete set of non-zero, outcome measures;

(4) 2,790 pupils who in addition had a complete set of covariates; and,

(5) 2,755 pupils who in addition were at sites with at least "five pupils per school."

C. Variables: Comparisons are based on five measures:

1. Spring Kindergarten (1972) Peabody Picture Vocabulary Test Scores,

2. Spring Kindergarten MAT-reading scores,

3. Spring Kindergarten MAT-numbers scores,

4. Spring First Grade (1973) MAT-total reading scores, and,

5. Spring First Grade MAT-mathematics scores.

D. Analytical Method: Within sponsors, scores for FT children in each of the five samples were compared with FT children in an undefined "random sample." Similarly, scores for NFT children in each of the five samples were compared with NFT children in an undefined "random sample." Differences greater than two standard errors and greater than one-quarter of a standard deviation are noted.

Tabular layout of means, standard deviations, and sample sizes permits visual comparison of five groups and provides the basis for calculating the significance of intergroup differences.

E. Comments:

1. Comparing each group with an unidentified reference group,

as opposed to comparing each group with other groups, is a very indirect way of approaching the research question.

2. All comparisons are conducted within sponsors. Results cannot be generalized beyond that level of analysis.

F. Results: All the specific comparisons are too numerous to mention. Readers are referred to the tables on pp. II-16 to II-20. The general conclusion drawn was: "The restrictions employed in the sample definition process have resulted in differential attrition for the Sponsor FT/NFT groups. The fact that differential attrition has occurred at the Sponsor level (some affected, others not), at the treatment level within Sponsor (in some Sponsors both the FT and the NFT; in others only the FT or NFT), and at the test level (some tests for some groups, other tests for other groups) severely limits our ability to both generalize to larger FT/NFT groups within Sponsors and to compare Sponsor results." (pp. II-21 - II-22).

"Comparison of 1974 Analytic Sample with Fall 1971 Tested Population and [the] 1973 Analytic Sample." (pp. II-22 - II-38)

A. Research Question: Is the 1974 Analytic Sample used for the overall FT/NFT contrast and the sponsor effects analysis significantly different from the "Fall 1971 Tested Population" or from the "1973 Analytic Sample"?

B. Sample: Three groups are mentioned:

"Fall 1971 Tested Population" -- This group remains undefined. The sum of FT and NFT, PPVT scores across all ten Sponsors is 14,987 -- a figure that does not correspond to the size of any group discussed.

"1973 Analytic Sample" -- This group also remains undefined. The sum of FT and NFT, PPVT scores across all ten Sponsors is 4009 -- a figure that does not correspond to the size of any group discussed.

"1974 Analytic Sample" -- The size of this group (N = 2755) indicates that this is the group of children which remained after applying the full set of sampling restrictions. It is the group used to analyze one-year and two-year effects.

C. Variables: The mentioned groups were compared on the basis of four measures:

1. The fall 1971 Kindergarten WRAT Reading Subtest,
2. The fall 1971 Kindergarten WRAT Spelling Subtest,
3. Mother's education index, and,
4. Family income index.

D. Method of Analysis: Means of FT children in the 1974 Analytic Sample were compared with the means of FT children from a group of children tested Fall 1971 and with the means of children from the 1973 Analytic Sample. The means of NFT children were compared in a similar way. Differences greater than two standard errors and greater than one-quarter standard deviation were noted.

E. Comments: Since the two samples with which the 1974 Analytic Sample is compared remained undefined -- i.e., how each was determined is not discussed -- the meaning of the reported differences is not clear.

It is worth noting that although the discussion in "Sample Development" purports to deal with the "1974 Analytic Sample", no discussion of the representativeness of the sample underlying the by-far greater part of the analyses of both Vols. II A and

II B is undertaken. Most of Vols. II A and B have to do with FT/NFT comparisons within Sponsors at the site level. A sample of 3481 children (2168 FT/1313 NFT) was selected for this purpose using different sampling restrictions (pp. II-5, 8, 9). The characteristics of this sample, its representativeness vis-a-vis the initial population of tested 1971 EK's, the degree of similarity between it and the "1974 Analytic Sample" remain undiscussed.

The "1974 Analytic Sample" used only on analysis of one-year and two-year, within-Sponsor FT/NFT contrasts differs from the "By-Site Sample" in at least one important aspect. The "1974 Analytic Sample" includes 47 sites; the "By-Site Sample" includes only 37. Although both samples are derived from a common subset of 4625 pupils who had a complete F71/S72/S73 battery of tests, had remained in first grade 72/73, and had stayed with the same Sponsor, there is no indication of the degree to which the 1724 FT of the 1974 Analytic Sample are among the 2,168 FT of the "By-Site Sample." The same could be said about the NFT pupils in both samples. In sum, it is impossible to say that the two samples are similar. Consequently, the discussion in "Sample Development" has little to do with the representativeness of the sample used for most of the subsequently reported effects analysis.

F. Results: Particulars are reported on pp. II-23 to II-38 and are too numerous to mention. In general:

"...it appears that there are differences among and within Sponsors on the representativeness of the FT/NFT groups in the current [read 1974 Analytic Sample] sample." (p. II-26)

"The current samples for several Sponsors differ not only from the 1971 Fall tested populations but also from last year's analytic samples as well." (p. II-38)

Moreover, with reference to Sponsor by Sponsor comparisons:

"The Sponsors' samples differ from each other in all possible ways. Sponsors vary in the extent to which the sample represents the population as well as the magnitude, direction and type of FT/NFT mismatch. The Sponsors are working with different children from different families in different locations under different circumstances. Since the meaning of any Sponsor effect is inextricably tied to unique characteristics of the sample on which it is based, we cannot compare effects across Sponsors." (p. II-39)

II. ONE YEAR, KINDERGARTEN STUDY* (Vol. II A, pp. V-1 - V-29)

A. Research Question: Does one year of kindergarten experience in a Follow Through classroom result in discernable outcomes?

B. Sample: Cohort III, EK children were used to determine whether there is an overall Sponsor effect, i.e., a specific Sponsor effect across all sites for a given Sponsor -- the ANCOVA Analytic Sample of 2755 was used.

		10 Sponsors										
		2	3	5	7	8	9	10	11	12	14	
Children:	FT	219	260	119	95	320	168	173	150	153	67	1724
	NFT	175	166	67	82	113	145	74	96	60	50	1031
												2755
Sites:		6	6	4	4	6	6	5	5	2	3	47

C. Variables:

1. Independent: exposure to FT or NFT experience during kindergarten.

2. Dependent: Spring 1972 scores on WRAT-spelling, WRAT-reading, WRAT-arithmetic, MAT-listening, MAT-reading, MAT-number, PPVT, Gumpgookies, Locus of Control-positive, Locus of Control-negative.

3. Covariates: Two sets were used -- one for Sponsor effects summed across all sites for a given Sponsor -- i.e.,

---

*This analysis appears in the Abt report under the title: "Kindergarten Study: A Replication" (p. V-1 ff.). We have changed the reference since the analysis is a replication only in a broad sense of the term. The unit of analysis is the child whereas the study being "replicated" -- i.e., the "Sponsor Vignettes" of Cline et al. (3/74) -- used school, class and child. The number of covariates used for the child level analysis has been reduced from ten to three for the overall Sponsor effect and two for the site level effect, and two or three for family income. Also, different outcome measures are used in this study: (1) S72 WRAT scores for three subtests (spelling, reading and arithmetic) are used instead of the single Spring 72 WRAT score and (2) absence is dropped as an outcome measure.

F71 WRAT, Mother's Education, and Family Income*; another for Sponsor effect at the site level -- i.e., F71 WRAT, Mother's Education, and presence/absence of mother's education score.

D. Analytical Method: The child is the unit of analysis. Children's scores are summed across sites within each Sponsor. FT/NFT contrasts within each Sponsor are examined using ANOVA and ANCOVA models.

E. Comments: Evidence indicates very unequal attrition of students across Sponsors and sites. "The Magnitude of the loss for each Sponsor must be considered as a reflection of the unique factors at each site. Unfortunately, these factors cannot be identified or quantified, and must remain as a source of confounding of Sponsor effects." (p. V-2)

F. Results: Adjusted, within-Sponsor effects are reported on the following page.

---

*The text states that "mother's education and entry MAT inequalities" were used (p. V-5). Elsewhere, in the discussion of covariates (p. IV-11) and in the discussion of Methodological Issues (Appendix B, p. B-24) the three covariates listed here are mentioned.

Rough Summary of Adjusted* FT/NFT Contrasts Within Sponsors

at End of K S'72**--Abt II A

Sponsors	Outcomes WRAT Spell	WRAT Read	WRAT Arith.	MAT Listen	MAT Read	MAT Numbers	PPVT	Gump-Gookies
2. Far West Labs	NFT	FT	---	---	FT	---	---	---
3. U. of Arizona	---	FT	NFT	NFT	---	NFT	---	---
5. Bank Street	NFT	NFT	---	NFT	NFT	---	---	---
7. U. of Oregon	FT	FT	FT	---	FT	FT	---	---
8. U. of Kansas	FT	FT	FT	---	FT	FT	FT	FT
9. High/Scope	---	FT	---	---	FT	---	---	---
10. U. of Florida	NFT	---	---	---	---	NFT	---	NFT
11. EDC	NFT	---	NFT	---	---	---	---	NFT
12. U. of Pitt.	---	NFT	FT	---	---	FT	---	---
14. SEDL	NFT	---	NFT	---	FT	---	---	---

*FT'71 WRAT, Mother's Education, and Family Income are the covariates.

**This is a rough summary, nothing more. Sponsor to sponsor comparisons are hazardous on several grounds, the most important of which are within Sponsor, inter-site variability and inadequacy of ANCOVA adjustments.

"FT" entry indicates that the adjusted FT/NFT difference was "statistically significant" at the .05 level, favoring FT.

"NFT" indicates that the adjusted difference was "statistically significant," favoring the NFT comparison group.

Locus Posit.	Locus Neg.
---	---
NFT	---
---	---
FT	---
---	---
FT	---
---	---
---	NFT
---	---
FT	---

III. TWO-YEAR EFFECTS (Vol. II A, pp. VI-1 - VI-126)

A. Research Question: Does two years of FT experience (kindergarten and first grade) have a demonstrable effect?

B. Sample: Two groups were used for two different analyses:

1. To determine whether there is an overall sponsor effect, the 1974 Analytic Sample of 2,755 children was used. (See preceding summary for a description of the distribution of children across sponsors.)

2. To determine Sponsor effects at the site level a "By-Site Sample" of 3,481 children was used. The distribution of children across sites and sponsors was as follows:

		10 Sponsors										
		2	3	5	7	8	9	10	11	12	14	
Children:	FT	263	367	102	132	332	240	270	138	184	140	2168
	NFT	187	242	78	100	131	202	141	84	64	84	1313
												3481
Sites:		5	5	2	5	4	6	4	3	3	3	37

C. Variables:

1. Independent: Two years of FT or two years of NFT experience.

2. Dependent: Only three outcome measures were employed:

   a. MAT-word analysis;

   b. MAT-mathematics;

   c. MAT-total reading.

3. Covariates: For the first analysis -- i.e., sponsor effects summed across all sites -- a single covariate was used -- Fall 1971 WRAT score. For the site level analysis of sponsor effects, three covariates were used:

   a. Fall 1971 WRAT;

   b. Mother's education;

   c. Presence or absence of Mother's education score.

D. Analytical Method: The child is the unit of analysis. Children's scores are summarized across sites within each sponsor. FT/NFT contrasts within each sponsor are examined using ANOVA and ANCOVA models.

E. Comments:

1. Reported sponsor effects (or non-effects) summed across all sites for a given sponsor are very difficult to interpret for a variety of reason. Three are mentioned in this study:

   a. "Sites are very different in respect to the administration and operation of FT programs and in the characteristics of the comparison NFT groups. Combining sites means combining diverse events, which tends to obscure important processes."(p. VI-3) We would add that it also confounds the analysis of the FT/NFT main effect.

   b. The application of sampling restrictions appears to have led to differential "attrition." That is, different sorts of children were lost from the FT and NFT groups and different sorts of children were lost at different sites. The analysis of FT/NFT main effects is confounded by this differential attrition.

   c. When sponsor effects are summed across sites, the adjustment for differences due to F71 WRAT entry scores is not comparable for both the FT and NFT groups -- i.e., there is some serious heterogeneity of regression.

2. The use of a single covariate -- F71 WRAT scores -- seems uneqal to the task of adjusting for considerable mismatch between FT and NFT groups. The meaning of the reported adjusted differences is difficult to determine.

F. Results: The results are summarized in the two following tables.

Summary of Within-Sponsor FT/NFT Contrasts at the End of Two Years of FT Experience*

Sponsors	Outcomes MAT Word Analysis	 MAT Mathematics	 MAT Total Reading
2. Far West Labs			
unadjusted	NFT	NFT	NFT
adjusted[a]	---	---	b
3. U. of Arizona			
unadjusted	NFT	NFT	NFT
adjusted	NFT	NFT	NFT
5. Bank Street			
unadjusted	NFT	---	NFT
adjusted	NFT	NFT	b
7. U. of Oregeon			
unadjusted	---	NFT	FT
adjusted	---	NFT	---
8. U. of Kansas			
unadjusted	FT	FT	---
adjusted	FT	FT	---
9. High/Scope			
unadjusted	NFT	NFT	NFT
adjusted	NFT	NFT	NFT
1[illegible]. U. of Florida			
unadjusted	NFT	---	NFT
adjusted	NFT	b	NFT
11. EDC			
unadjusted	NFT	FT[c]	---
adjusted	NFT	FT[c]	---
12. U. of Pittsburgh			
unadjusted	---	---	---
adjusted	---	b	b
13. SEDL			
unadjusted	---	---	FT
adjusted	---	b	b

---

*This is a rough summary, nothing more. Sponsor to sponsor comparisons are

---

precluded on several grounds, the most important of which are within-Sponsor, inter-site variability and inadequate ANCOVA adjustments for initial differences. See subsequent pages for summary of site-by-site comparisons.

"FT" entry indicates that the FT/NFT difference was "statistically significant" at the .05 level, favoring FT.

"NFT" indicates that the difference was "statistically significant," favoring the NFT comparison group.

"---" indicates that the FT/NFT difference was not "statistically significant."

a/ Adjusted using F71 WRAT scores.

b/ Heterogeneity of regression relations precluded covariance adjustments.

c/ These comparisons were not marked as statistically significant in Abt's report (p. VI-98), but large F-ratios indicate that the comparisons were in fact statistically significant.

Summary of Site-Level Analyses of Two-Year Effects*

	Outcomes		
2. FAR WEST LABS (5 Sites)	MAT Word Analysis	MAT Mathematics	MAT Total Reading
1) NC-1 Duluth			
unadjusted	FT	---	FT
adjusted[a]	FT	---	FT
2) NC-2 St. Louis			
unadjusted	FT	FT	FT
adjusted	FT	FT	FT
3) NE Buffalo			
unadjusted	---	---	---
adjusted	---	---	---
4) W-1 Fresno, CA			
unadjusted	NFT	NFT	NFT
adjusted	NFT	NFT	NFT
5) W-2 Salt Lake City			
unadjusted	---	NFT	---
adjusted	---	---	---
3. UNIVERSITY OF ARIZONA (5 Sites)			
1) S Chickasha, OK			
unadjusted	NFT	NFT	NFT
adjusted	NFT	NFT	NFT
2) NC-1 Lincoln, NE			
unadjusted	NFT	NFT	NFT
adjusted	NFT	NFT	NFT
3) NC-2 Wichita, KA			
unadjusted	NFT	NFT	NFT
adjusted	NFT	NFT	NFT
4) NE-1 Lakewood, NJ			
unadjusted	NFT	NFT	NFT
adjusted	NFT	NFT	NFT
5) NE-2 Newark, NJ			
unadjusted	---	---	NFT
adjusted	---	NFT	NFT

	Outomes		
5. BANK STREET (2 Sites)	MAT Word Analysis	MAT Mathematics	MAT Total Reading
1) NE-1 Brattleboro, VT			
unadjusted	NFT	NFT	NFT
adjusted	NFT	NFT	NFT
2) NE-2 Fall River, MA			
unadjusted	NFT	—	NFT
adjusted	NFT	—	NFT
7. UNIVERSITY OF OREGON (3 Sites)			
1) NC-1 Grand Rapids, MI			
unadjusted	---	---	FT
adjusted	---	NFT	---
2) NC-2 Flint, MI			
unadjusted	FT	---	FT
adjusted	---	---	FT
3) NE Providence, RI			
unadjusted	---	---	---
adjusted	---	---	---
8. UNIVERSITY OF KANSAS (4 Sites)			
1) NE Trenton, NJ			
unadjusted	---	FT	---
adjusted	---	FT	---
2) NC-2 Kansas City, MO			
unadjusted	---	FT	---
adjusted	FT	FT	FT
3) S Louisville, KY			
unadjusted	FT	FT	FT
adjusted	FT	FT	FT
4) NC-1 Portageville, MO			
unadjusted	---	---	---
adjusted	---	---	NFT

	Outcomes		
9. High/Scope (6 Sites)	MAT Word Analysis	MAT Mathematics	MAT Total Reading
1) NE New York, NY			
unadjusted	---	---	---
adjusted	---	---	---
2) NC Chicago, IL			
unadjusted	---	---	NFT
adjusted	---	---	NFT
3) W-1 Greeley, CO			
unadjusted	---	---	---
adjusted	---	---	---
4) W-2 Denver, CO			
unadjusted	---	---	---
adjusted	FT	---	---
5) S Fort Walton Beach/Okaloosa County, FL			
unadjusted	NFT	NFT	NFT
adjusted	NFT	NFT	NFT
6) W-3 Seattle, WA			
unadjusted	NFT	NFT	NFT
adjusted	--	NFT	NFT
10. UNIVERSITY OF FLORIDA (4 Sites)			
1) S-1 Richmond, VA			
unadjusted	---	---	---
adjusted	---	---	---
2) S-2 Houston, TX			
unadjusted	FT	---	FT
adjusted	FT	---	---
3) W Yakima, WA			
unadjusted	---	---	FT
adjusted	---	---	---
4) NC Lawrenceburg, IN			
unadjusted	---	NFT	NFT
adjusted	---	---	---

11. EDC (3 Sites)	Outcomes		
	MAT Word Analysis	MAT Mathematics	MAT Total Reading
1) NE-1 Scranton/Lackawanna County, FL			
unadjusted	---	---	---
adjusted	---	FT	---
2) NE-2 Patterson, NJ			
unadjusted	NFT	FT	NFT
adjusted	NFT	FT	NFT
3) S Washington/Morgan Community School, D.C.			
unadjusted	---	FT	---
adjusted	FT	FT	FT
12. UNIVERSITY OF PITTSBURGH (2 Sites)			
1) NC Montevideo, MN			
unadjusted	---	---	---
adjusted	---	---	---
2) NE Lock Haven, PA			
unadjusted	NFT	NFT	NFT
adjusted	NFT	---	NFT
14. SEDL (3 Sites)			
1) NE Philadelphia, PA			
unadjusted	---	---	FT
adjusted	---	---	FT
2) W Tulare, CA			
unadjusted	---	FT	FT
adjusted	FT	FT	FT
3) S St. Martin Parishville, LA			
unadjusted	---	---	---
adjusted	---	---	---

---

*Site level analysis was undertaken for several reasons, one of which was between site variability on a number of factors. The findings reported here become meaningful only when they are considered in light of the characteristics of FT and NFT groups at each site. Only then can the questions of which Sponsor model works best with which sort of child in which type of community begin to be addressed. Only then can reasons for FT effects or noneffects begin to be puzzled out. The site to site FT to NFT comparisons on these characteristics are too numerous to mention. Readers are referred to the pertinent pages in Abt II A, Chapter 6.

"FT" entry indicates that the FT/NFT difference was "statistically significant" at the .05 level, favoring FT.

"NFT" indicates that the difference was "statistically significant," favoring the NFT comparison group.

"---" indicates the FT/NFT difference was not "statistically significant."

Notes:

a. Adjusted for F71 WRAT.

IV. INTERIM ANALYSIS OF LONGITUDINAL FT/NFT SITE LEVEL DIFFERENCES IN ACADEMIC ACHIEVEMENT (Vol. II B, pp. I-1 - I-89)

The previous two summaries have described efforts to estimate 1) the effects of one year FT experience (kindergarten) and 2) the accumulated effect of two years in FT (kindergarten plus the first grade). The analysis summarized here attempted to go one step further and separate out the effects that result exclusively from the first grade FT experience.

A. Research Questions: "What are the differences between gains in FT and NFT groups over the kindergarten and first grade period?

- which are attributable to kindergarten experience?

- which are attributable to first grade experience?

- which are attributable to differences between FT and NFT at entry into kindergarten?

"Do these differences between gains over the two years suggest any trends?

"Do children, FT or NFT, who start with a disadvantage at entry catch up to and/or surpass their comparison group? Do they

- maintain their entry status with respect to their comparison?

- fall progressively further behind their comparison?

"Do sponsor's FT groups show consistent/inconsistent patterns across grades? Across sites?" (p. I-1)

B. Sample: 3481 children, distributed across ten sponsors and 37 sites in the following manner:

		10 Sponsors									
		2	3	5	7	8	9	10	11	12	14	
Children:	FT	263	367	102	132	332	240	270	138	184	140	21
	NFT	187	242	78	100	131	202	141	84	64	84	131 34
Sites:		5	5	2	3	4	6	4	3	2	3	

C. Variables:

1. Independent: FT or NFT experience.

2. Dependent: FT/NFT differences were estimated in terms of the test scores from test administrations spring 1973.

   a. MAT-word analysis.

   b. MAT-mathematics.

   c. MAT-total reading.

D. Analytical Method: ANOVA and ANCOVA models based on multiple regression effects coding were used to generate three estimates of FT and NFT differences. The first was the unadjusted FT/NFT differences in Spring 1973 scores -- i.e., the regression coefficient b_1. The second was the FT/NFT differences on Spring 1973 scores adjusted for differences on Fall 1971 entry kindergarten (presumably WRAT scores) -- b_2. And the third estimated difference was the FT/NFT difference on Spring 1973 scores adjusted for both differences on fall 1971 WRAT scores and differences in exit from kindergarten (presumably WRAT) scores -- b_3.

These three effects estimates were used to calculate FT/NFT differences attributable to experiences in three time periods:

a) experience previous to entering kindergarten,

b) experience during kindergarten; and

c) experience during the time between exit from kindergarten and exit from first grade.

The following three time period related differences were calculated:

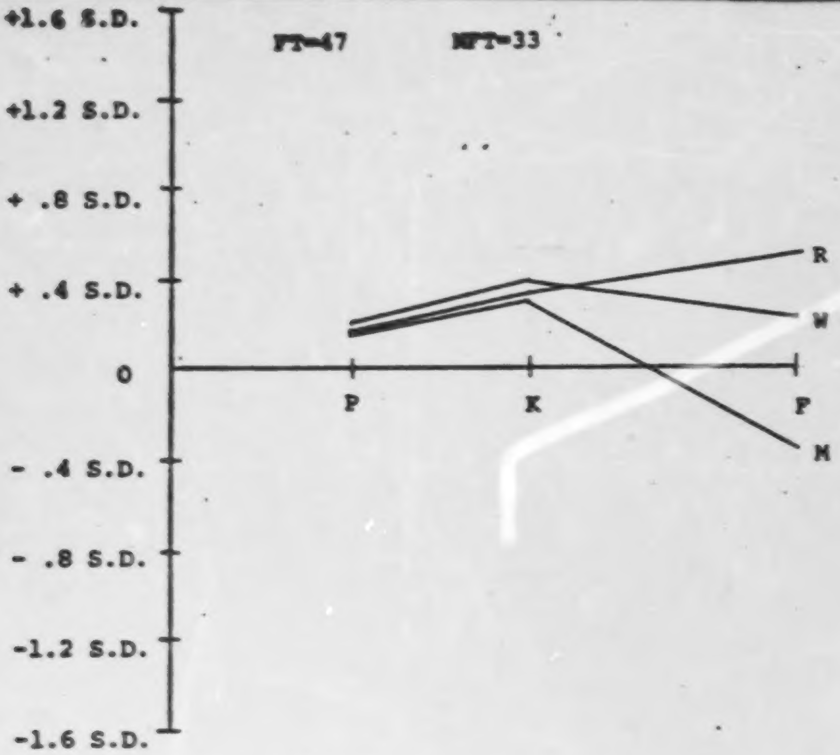
F -- (b_1) interpreted as the FT/NFT differences at the end of the first grade,

K -- (b_1-b_3) interpreted as "the expected difference on the first grade outcome based on the combined differences on kindergarten entry and exit tests." (p. I-11)

P -- (b_1-b_2) interpreted as "the expected difference on the first grade outcome based on differences at kindergarten entry." (p. I-10)

The results of the analysis are reported in the manner shown on the following page. The ANOVA based "Unadjusted Difference" (b_1) is reported in the table at the lower half of the figure. Also reported in the table is the accumulated "Two-Year

Difference", adjusted for entry kindergarten scores (b_2). And the accumulated FT/NFT differences adjusted for both entry and exit from kindergarten scores are reported as "First Grade Difference". Differences greater than two standard errors are noted.



	MAT WORD ANALYSIS	MAT MATH	MAT READING
Unadjusted Difference (A)	.0495	- 3.9400	8.1103*
S.E.	.0481	2.4239	3.4544
Two-Year Difference (B)	.0080	- 5.4614*	5.6480
S.E.	.0433	2.2306	2.9360
First Grade Difference (C)	- .0345	- 7.2317*	2.7648
S.E.	.0412	2.1961	2.7933
Outcome S.D.	.2120	10.7829	15.6387
P=(A-B)/S.D.	.1958	.1411	.1574
K=(A-C)/S.D.	.3962	.3053	.3418
F=A/S.D.	.2335	- .3654	.5186

Figure I-13 .. First Grade Outcome Profile for Oregon Site NC1

Source: Volume B, p. I-35.

Note: All items marked by * are significant ($p \leq .05$).

In addition to the tabular display of unadjusted and adjusted effects estimates, the report "graphs" P, K, and F values in the manner indicated at the top of the figure.

The child is the unit of analysis.

FT/NFT differences are estimated at the site level.

E. Comments:

1. Sample: Although the report implies that the sample used is the 1974 Analytic Sample (N = 2755)*, the sum of reported site level samples for each sponsor indicates that the sample is the By-Site Sample of 3,481 children.

2. Variables:

   a. Readers should note that one outcome measure, MAT-word analysis, is a subcomponent of another, MAT-reading.

   b. The report does not specify which entry and exit kindergarten outcomes are used as covariates. We have presumed that WRAT scores were used, but that is a presumption.

   c. Coefficients of internal consistency for each of the three outcome measures ranged from .93 to .96.

3. Analytical Method: In attempting to explain its methodology, the report states:

   "The results are presented in charts which show differences between FT and NFT; FT-favoring differences are shown by upward sloping lines, and NFT-favoring differences by downward sloping lines. The figures do not show gains or growth. Rather they show differences in gains between FT and NFT as these differences are related to subsequent differences in such gains." (p. I-5, emphasis in the original.) The report goes on to say that the P, K, and F values described above are entered into figures. "The figure shows the differences between FT and NFT at entry to kindergarten, at exit from kindergarten, and at exit from first grade in standard deviation units of the outcome measures. The distance between the points is intended to reflect the time interval between testing: approximately 3/4 of a year for the kindergarten entry to exit interval approximately one year for the kindergarten exit to first grade exit interval." (p. I-5) Yet, "Although these figures present longitudinal patterns of differences across kindergarten and first grade, they do not represent achievement domains measured across time." (p. I-7, emphasis in the original.) We had some difficulty in understanding these conflicting explanations and resorted to our own interpretations of the report's intent in calculating the P, K, and F values.

---

*The report states on p. I-5 that the sample used in the analysis is described in Appendix A of Volume IIA of the report. The sample described in this Appendix is the 1974 Analytic Sample.

The calculation of P, K, and F values is the direct result of students not being tested with the same examinations at different time points. In the absence of such measures, efforts were made to estimate how students might have differed on the three spring 1973 outcome measures had they taken them spring 1972 and when they entered kindergarten fall 1971. Another way of looking at these estimates is to consider them as retrospective projections or "retrojections". That is, instead of using multiple regression techniques to predict differences in the future, these techniques are used to estimate how scores would have been different in the past. There is an inherent logic to this approach, but the approach has certain difficulties which need to be noted. Moreover, the manner in which the analyses were conducted is open to question. And, the way in which the results were reported can be quite misleading.

As far as the approach goes, its underlying logic seems quite extenuated. Its application seems prone to generate results that are difficult to relate to the initial research questions. To say that P values indicate how students would have scored differently on a test appropriate for testing skills at the end of the first grade, had they taken it when they entered kindergarten 21 months earlier, is somewhat of a convoluted approach to the matter of early kindergarten differences.

At a more technical level, the calculation of adjusted effects estimates (both b_2 and b_3) assumes that a single academic achievement measure when used as a covariate is adequate to the task of adjusting for all important FT/NFT differences at the beginning of respective time periods. Such an assumption is highly questionable.

Another technical note to be made concerning the procedures used is a matter of factor orderings (the sequence in which variables in the ANOVA and ANCOVA models are entered into the regression analysis). The unequal number of FT and

NFT students is alone cause enough to ensure that the contrast designs are not orthogonal. The report makes no mention of having tested alternative factor orderings to determine whether significantly different effects estimates would have been obtained with factor orderings different from the ordering apparently used.

The most serious problem with the procedures followed in the present analyses is the manner in which P, K, and F values are "graphed." The particular symbols used are not neutral. They imply points in time -- i.e., P for pre-school, K for kindergarten, and F for first grade. Laying them out sequentially on what appears to be a conventional graph clearly implies that the lines connecting the P, K, and F points indicate a longitudinal pattern in the scores. Yet such an interpretation is not warranted. The vertical distance between points K and F represents a difference between adjusted and unadjusted FT/NFT differences -- not change in scores over time. Similarly, the graphed vertical distance between P and K represents a difference between (1) FT and NFT scores adjusted for both entry and exit kindergarten scores -- not FT/NFT differences at entry to kindergarten as opposed to differences at exit from kindergarten. In short, both the explicit and implicit causal and/or temporal interpretations given to the results are very inappropriate.

F. Results: Since unadjusted effects estimates and effects estimates adjusted for entry kindergarten scores have already been summarized earlier when presenting the results of the Two-Year Effects analysis the present summary focuses exclusively on the results having to do with effects estimates adjusted for both entry and exit kindergarten scores. The site by site FT/NFT differences for the respective sponsors are summarized in the following table.

Rough Summary of Two-Year Effects: A Comparison of Unadjusted, Adjusted for Entry Kindergarten, and Adjusted for Both Entry Kindergarten and Exit Kindergarten Results *

2. FAR WEST LABS 5 Sites	Outcomes MAT Word Analysis	 MAT Mathematics	 MAT Total Reading
1) NC-1 Duluth			
unadjusted	FT	FT	FT
adjusted for F71 WRAT	FT	---	FT
adjusted for F71 & S72 WRAT	FT	---	FT
2) NC-2 St. Louis			
unadjusted	FT	FT	FT
adjusted for F71 WRAT	FT	FT	FT
adjusted for F71 & S72 WRAT	FT	---	---
3) NE Buffalo			
unadjusted	---	---	---
adjusted for F71 WRAT	---	---	---
adjusted for F71 & S72 WRAT	---	---	---
4) W-1 Fresno, CA			
unadjusted	NFT	NFT	NFT
adjusted for F71 WRAT	NFT	NFT	NFT
adjusted for F71 & S72 WRAT	NFT	NFT	NFT
5) W-2 Salt Lake City			
unadjusted	---	NFT	---
adjusted for F71 WRAT	---	---	---
adjusted for F71 & S72 WRAT	---	NFT	NFT

---

* The findings reported here become meaningful only when they are considered in light of the characteristics of FT and NFT groups at each site. Only then can the questions of which Sponsor model works best with which sort of child in which type of community begin to be addressed. Only then, can reasons for FT effects or noneffects begin to be puzzled out. The site to site FT to NFT comparisons on these characteristics are too numerous to mention.

"FT" entry indicates that the FT/NFT difference was "statistically significant" at the .05 level, favoring FT.

"NFT" indicates that the difference was "statistically significant", favoring the NFT comparison group.

"---" indicates that the FT/NFT difference was not statistically significant."

	Outcomes		
3. UNIVERSITY OF ARIZONA 5 Sites	MAT Word Analysis	MAT Mathematics	MAT Total Reading
1) S Chickasha, OK			
unadjusted	NFT	NFT	NFT
adjusted for F71 WRAT	NFT	NFT	NFT
adjusted for F71 & S72 WRAT	NFT	NFT	NFT
2) NC-1 Lincoln, NE			
unadjusted	NFT	NFT	NFT
adjusted for F71 WRAT	NFT	NFT	NFT
adjusted for F71 & S72 WRAT	NFT	NFT	NFT
3) NC-2 Wichita, KA			
unadjusted	NFT	NFT	NFT
adjusted for F71 WRAT	NFT	NFT	NFT
adjusted for F71 & S72 WRAT	NFT	NFT	NFT
4) NE-1 Lakewood, NJ			
unadjusted	NFT	NFT	NFT
adjusted for F71 WRAT	NFT	NFT	NFT
adjusted for F71 & S72 WRAT	NFT	NFT	NFT
5) NE-2 Newark, NJ			
unadjusted	---	---	NFT
adjusted for F71 WRAT	---	NFT	NFT
adjusted for F71 & S72 WRAT	---	NFT	NFT
5. BANK STREET 2 Sites			
1) NE-1 Brattleboro, VT			
unadjusted	NFT	NFT	NFT
adjusted for F71 WRAT	NFT	NFT	NFT
adjusted for F71 & S72 WRAT	NFT	NFT	NFT
2) NE-2 Fall River, MA			
unadjusted	NFT	---	NFT
adjusted for F71 WRAT	NFT	---	NFT
adjusted for F71 & S72 WRAT	---	FT	NFT
7. UNIVERSITY OF OREGON 3 Sites			
1) NC-1 Grant Rapids. MI			
unadjusted	---	---	FT
adjusted for F71 WRAT	---	NFT	---
adjusted for F71 & S72 WRAT	---	NFT	---
2) NC-2 Flint, MI			
unadjusted	FT	---	FT
adjusted for F71 WRAT	---	---	FT
adjusted for F71 & S72 WRAT	---	NFT	---
3) NE Providence, RI			
unadjusted	---	---	---
adjusted for F71 WRAT	---	---	---
adjusted for F71 & S72 WRAT	---	---	---

	Outcomes		
8. UNIVERSITY OF KANSAS 4 Sites	MAT Word Analysis	MAT Mathematics	MAT Total Reading
1) NE Trenton, NJ			
unadjusted	---	FT	---
adjusted for F71 WRAT	---	FT	---
adjusted for F71 & S72 WRAT	---	---	---
2) NC-2 Kansas City, MO			
unadjusted	---	FT	---
adjusted for F71 WRAT	FT	FT	FT
adjusted for F71 & S72 WRAT	---	---	---
3) S Louisville, KY			
unadjusted	FT	FT	FT
adjusted for F71 WRAT	FT	FT	FT
adjusted for F71 & S72 WRAT	---	---	---
4) NC-1 Portageville, MO			
unadjusted	---	---	---
adjusted for F71 WRAT	---	---	NFT
adjusted for F71 & S72 WRAT	---	---	NFT
9. HIGH/SCOPE 6 Sites			
1) NE New York, NY			
unadjusted	---	---	---
adjusted for F71 WRAT	---	---	---
adjusted for F71 & S72 WRAT	---	---	---
2) NC Chicago, IL			
unadjusted	---	---	NFT
adjusted for F71 WRAT	---	---	NFT
adjusted for F71 & S72 WRAT	---	---	NFT
3) W-1 Greeley, CO			
unadjusted	---	---	---
adjusted for F71 WRAT	---	---	---
adjusted for F71 & S72 WRAT	---	---	---
4) W-2 Denver, CO			
unadjusted	---	---	---
adjusted for F71 WRAT	FT*	---	---
adjusted for F71 & S72 WRAT	---	---	---

---

*The FT/NFT contrast is reported as insignificant in Volume B (p. I-54) although it is reported as significant in Volume A (p. VI-81). To be consistent with previously summarized results we have reported the difference as significant.

9. High/Scope (continued)	Outcomes		
	MAT Word Analysis	MAT Mathematics	MAT Total Reading
5) S Fort W Beach/Okaloosa County, FLA			
unadjusted	NFT	NFT	NFT
adjusted for F71 WRAT	NFT	NFT	NFT
adjusted for F71 & S72 WRAT	NFT	NFT	NFT
6) W-3 Seattle, WA			
unadjusted	NFT	NFT	NFT
adjusted for F71 WRAT	---	NFT	NFT
adjusted for F71 & S72 WRAT	NFT	NFT	NFT
10. UNIVERSITY OF FLORIDA 4 Sites			
1) S-1 Richmond, VA			
unadjusted	---	---	---
adjusted for F71 WRAT	---	---	---
adjusted for F71 & S72 WRAT	---	FT	---
2) S-2 Houston, TX			
unadjusted	FT	---	FT
adjusted for F71 WRAT	FT	---	---
adjusted for F71 & S72 WRAT	FT	---	---
3) W Yakima, WA			
unadjusted	---	---	FT
adjusted for F71 WRAT	---	---	---
adjusted for F71 & S72 WRAT	---	---	---
4) NC Lawrenceburg, IND			
unadjusted	---	NFT	NFT
adjusted for F71 WRAT	---	---	---
adjusted for F71 & S72 WRAT	---	---	---
11. EDC 3 Sites			
1) NE-1 Scranton/Lackawanna County, PA			
unadjusted	---	---	---
adjusted for F71 WRAT	---	FT	---
adjusted for F71 & S72 WRAT	---	FT	---
2) NE-2 Patterson, NJ			
unadjusted	NFT	FT	NFT
adjusted for F71 WRAT	NFT	FT	NFT
adjusted for F71 & S72 WRAT	NFT	FT	NFT
3) S Washington/Morgan Community School, D.C			
unadjusted	---	FT	---
adjusted for F71 WRAT	FT	FT	FT
adjusted for F71 & S72 WRAT	---	FT	---

	Outcomes		
12. UNIVERSITY OF PITTSBURGH 2 Sites	MAT Word Analysis	MAT Mathematics	MAT Total Reading
1) NC Montevideo, MINN			
unadjusted	---	---	---
adjusted for F71 WRAT	---	---	---
adjusted for F71 & S72 WRAT	---	---	---
2) NE Lock Haven, PA			
unadjusted	NFT	NFT	NFT
adjusted for F71 WRAT	NFT	---	NFT
adjusted for F71 & S72 WRAT	NFT	NFT	NFT
14. SEDL 3 Sites			
1) NE Philadelphia, PA			
unadjusted	---	---	FT
adjusted for F71 WRAT	---	---	FT
adjusted for F71 & S72 WRAT	---	FT	FT
2) W Tulare, CA			
unadjusted	---	FT	FT
adjusted for F71 WRAT	FT	FT	FT
adjusted for F71 & S72 WRAT	FT	FT	FT
3) S St. Martin Parishville, LA			
unadjusted	---	---	---
adjusted for F71 WRAT	---	---	---
adjusted for F71 & S72 WRAT	---	---	---

V. TEACHER STUDIES (Vol. II B, pp.II-1 - II-47)

A. Research Questions: "In order to describe more fully the dimensions of the Follow Through program impact, we have also examined the characteristics of the teachers who work with the children." (p. II-1) More specifically, this study is aimed at describing and comparing FT and NFT teachers with respect to:

-- personal and professional characteristics;

-- attitudes of FT teachers toward FT and sponsor;

-- focus of training received by teachers;

-- reported teacher goals and practices;

-- teachers' reports of parent involvement.

B. Sample: FT and NFT teachers responding to the spring 1973 teacher questionnaire (n = 1657). Various selection criteria (p. II-2) reduced this sample to 1430 (726 FT and 704 NFT). Because of missing data, sample sizes for comparing teacher background characteristics ranged slightly below this number, except for teacher ethnicity where n = 945.

No data are provided on the number of sites represented in these samples. The teachers came from CIII non-entering first grade, CIV kindergarten and other cohort groups (i.e., CIII second grade, CII second and third, CI third). (p. II-3 - II-7).

C. Variables:

1. Independent: Cohort, FT/NFT, Sponsors.

2. Dependent:

a. Personal and professional characteristics: age, total teaching experience with current school, experience with sponsors approach, education, certification, salary, and ethnicity.

b. Cluster variables: grouped as follows and with reliability coefficients (i.e., coefficients of internal consistency) in parentheses:

Amount of teacher training:

1) Social/Emotional/Intellectual Development and Individualization of Instruction (.93).

2) Classroom management (.88).

3) Parent Involvement (.90).

Teacher practices:

1) Emphasis on basic skills versus affective development (.64).

2) Flexible/Open classroom (.68).

3) Individualization of instruction (.51).

4) Teacher centeredness.

Importance of Parental Involvement:

1) In social/educational activities (.95).

2) In decision-making activities (.81).

3) In housekeeping activities (.74).

4) Benefits of parent involvement (.83).

Teacher Satisfaction with Pupils' Affective Development (.79).

Teacher Satisfaction with FT Sponsor (FT only) (.95).

Amount of Parent Advisory Council Activity (FT only) (.90).

c. Two individual items from TQ:

Teacher satisfaction with pupil progress in reading.

Teacher satisfaction with pupil progress in match and numbers concepts.

3. Covariates: Teacher education, years of elementary school experience and ethnicity were used as covariates in the ANCOVA applied to both cluster variables and the two single items.

D. Analytical Method: The means and standard deviations of teacher personal and professional characteristics variables are simply reported by cohort group, sponsor and FT/NFT (not by site).

Cluster variables were developed by organizing a "limited number of items" from the TQ into four groups and then applying cluster analysis to the grouped items for a random sampling of 400 teacher respondents (the report does not clearly report what larger sample this random sample was drawn from). (p. II-6)

Regression analysis using an ANCOVA design was applied to each of the cluster variables and the two single-item variables (using covariates mentioned above) in order to estimate sponsor and FT/NFT X sponsor differences ("effects") on these variables. (p. II-14)

E. Comments: The report does not mention response rate to the teacher questionnaire nor the problem of non-response bias. Nor does it indicate what information from the teacher questionnaire was not analyzed. It provides no indication of the variability of teacher responses by site within sponsor. Likewise, the report does not fully describe the ANCOVA design used in the study and fails to mention whether the assumptions of ANCOVA are even approximately met in this analysis.

The major weakness in the study, however, is that it attempts to infer impact of FT program on teachers from a post-test only evaluation design without even mentioning the problems implicit in doing so.

F. Results: "The Follow Through program does appear to be having a measurable impact on its teachers. They are receiving more training than their comparisons in social/emotional/intellectual development, classroom management, and parent involvement. The teachers practices...vary across Sponsors in terms of the use of an open/flexible classroom, individualization of instruction, and teacher-centeredness. These variations are, generally, consistent with each Sponsor's approach. FT also appears to be having a considerable impact on the use of parent involvement and the teachers' attitudes toward its use. These teachers' reports indicate that the parents' potential contributions are being sought and utilized to an increasing extent over that found in NFT groups. Finally, most FT teachers are satisfied with their Sponsor's model. This implies that the Sponsors have designed programs which are generally amenable to teachers' styles and pupils' needs for learning. While the match may not be perfect, more teachers appear to have adopted their Sponsor's approach to the education process.

"While we would have liked to have merged teacher and pupil data for this report, time and contractual limitations prevented us from doing so. While the teacher studies give us some confidence that we are in fact assessing the impact of different models we do not yet know the extent to which variations in implementation affect pupil outcomes.

"Age and Experience

"In general, these teachers are in their mid-thirties, have over nine years total experience, and have been teaching for over five years in their current schools. There are some variations by Sponsor and by cohort/grade across these variables. Further, while the FT teachers tend to be somewhat younger, have fewer total years of experience, and have been teaching fewer years in their current schools, they are not inexperienced teachers. In the case of the Cohort III teachers, the FT/NFT contrasts remain fairly consistent across Sponsors, with two exceptions. SEDL FT teachers are slightly older and have more experience both in teaching and in their current schools. Bank Street also tends to have FT teachers with more years of experience in their current schools." (p. II-14)

"While Sponsors vary somewhat in their teachers' level of experience with their approach, the Cohort III teachers, on the average, have slightly over two and one-half years of experience. EDC has the least experienced group, with an average of 1.85 years of experience with that approach; SEDL has the most experienced group, with an average of 3.28 years of experience."

"Education and Certification

"FT and NFT teachers differ very little in educational attainment. The cohort/grade groups show more variability: approximately 50 per cent or more, in contrast to 65 per cent of the teachers in the other cohort/grade groups. The Sponsor groups also vary considerably: among the Cohort III teachers, the NFT teachers tend to have more education than the FT teachers. In Kansas and High/Scope, however, more FT teachers than NFT have earned more than a Bachelor's degree. In the other cohort/grade groups, FT teachers tend to have a higher educational attainment level, but the Sponsors vary substantially." (p. II-20)

"The FT and NFT teacher groups differ in their levels of certification. For example, 71.4 per cent of the Cohort III NFT teachers are permanently certified, as opposed to 60.7 per cent of the FT teachers. In addition, the average percentages of FT and NFT teachers with renewable certification are 21.3 per cent and 19.4 per cent, respectively; and the average percentages of FT and NFT teachers with temporary certification are 18.0 per cent and 9.2 per cent respectively. There is considerable variability among Sponsors on these certification patterns, however.

"Bank Street and EDC have FT/NFT groups with over 70 per cent permanently certified teachers and approximately 20 per cent holding temporary certification. Fewer than eight per cent have renewable certification. A different pattern exists for Oregon and Florida. Both Sponsors are using approximately 50 per cent non-permanently certified FT teachers, two-thirds of whom have temporary certification, only. By comparison, approximately 65 per cent of the NFT teachers are permanently certified. In Arizona, High/Scope, and Pittsburgh, FT teachers tend to have either permanent or renewable certification. More NFT comparison teachers associated with these sponsors hold permanent certification. The distribution by region and city size of Sponsors with parallel certification patterns indicates that these patterns may be community related.

"In general, the FT/NFT sample distributions on certification indicate that those teachers who are not permanently certified, i.e., 'marginal teachers' are either assigned or attracted to the FT experiment. Even though the FT teachers tend to be experienced, and have attained a fairly high level of education, they tend to be somewhat marginal with respect to the teaching profession. This varies by Sponsor, but it does reveal FT's generally perceived experimental status." (p. II-23)

"Salary

"The NFT teachers tend to earn slightly more than the FT teachers. This is possibly a reflection of the different experience and certification levels of the two groups. This pattern is consistent across teachers for all Sponsors in the Cohort III group. The Cohort IV kindergarten teachers tend to be more highly paid, probably because they teach in the big cities, where salaries tend to be higher. These Cohort IV teachers also tend to be more experienced and have a higher degree of educational attainment.

"Ethnicity

"Overall, there are more minority teachers in the FT than in the NFT classes. In Cohort III, teacher ethnicity varies considerably among the Sponsors, paralleling, in some cases, that of the pupils. For example, Pittsburgh FT/NFT pupils and teachers are both over 98 per cent White , while SEDL pupils and teachers are both slightly less than 50 per cent White. In other instances, such as Oregon with 5.3 per cent White pupils vs. 71.4 per cent White teachers and 19.5 per cent White NFT pupils vs. 79.5 per cent White teachers, there is a considerable difference. Unfortunately, without merging teacher/pupil data, we do not know in what way teacher-pupil ethnic match, or any of the other teacher background variables, mediate Sponsor impact on pupil performance." (p. II-25.)

"Teacher Satisfaction with FT Sponsor

"In general, no Sponsor group differs greatly from the overall FT mean. There appears to be a general degree of teacher satisfaction with each of the models. There are, however, some differences among Sponsor groups which are worthy of note." (p. II-25)

"The teachers in Cohort III who are associated with Bank Street and Pittsburgh express the highest degree of satisfaction with their Sponsor's program. The teachers associated with Oregon express the least degree of satisfaction. It is interesting that teachers in two programs which emphasize a highly individualized application of teacher skill and training (Bank Street), and curriculum materials (Pittsburgh), indicate more satisfaction than those teachers (Oregon) who rely primarily on a structured curriculum designed for use with groups of children. Whether the teachers are expressing more satisfaction about the model suiting their style or suiting the needs of the pupils is not clear, since this variable is a combination of these and other facets associated with the model.

"The expressed satisfaction of the Cohort III teachers is not entirely consistent with the Cohort IV group or with the 'other' group...It is not possible to identify the basic reasons for the differences in expressed satisfaction. As was mentioned previously, these groups are not comparable. The distributions of teachers by region and city size are very different. The differences in the displayed results could be attributed to either demographic or grade level factors." (p. II-28)

"...teachers associated with Far West and Bank Street report receiving significantly more training in social/emotional/intellectual development and individualization of instruction than their respective NFT teachers. Only Florida and Pittsburgh FT teachers report receiving less training on this dimension than their NFT comparisons. Their NFT groups are receiving more training than the other NFT groups on this dimension.

"Arizona, Bank Street and Kansas place considerably more emphasis on this dimension of training. The FT teachers in Florida and EDC report receiving somewhat less training in classroom management than their NFT comparisons.

"The teachers associated with Bank Street, Kansas, Florida, EDC, and SEDL indicate they are receiving significantly more training in parent involvement than the NFT's for each Sponsor. It is noteworthy that no Sponsor's teachers report receiving less training in parent involvement than the NFT's.

"It is clear...that the Sponsors differ in both the amount of training and the content areas which receive emphasis. FT teachers in all Sponsor programs report receiving more training than the NFT teachers. The patterns reported by the teachers, however, fluctuate. In some cases the variations reflect Sponsor orientations. For example, Oregon, which places considerably more emphasis on using prescribed curriculum materials than on teacher training, has teachers reporting that they receive relatively little training. Florida has a heavy emphasis on parent involvement in the curriculum delivery; this is reflected in the teachers' reports on the relative amount of training placed on this component. In some cases they do not. The Pittsburgh FT teachers could be expected to show a positive contrast with the NFT teachers on the social/emotional/intellectual development and individualization of instruction component. Apparently, the items pertaining to individualization of instruction are overshadowed by the other variables entering the cluster." (p. II-29)

Reported Teacher Practices

Different sponsors' teachers reported no overall significant differences on the "Emphasis on Basic Skills vs. Affective Development" cluster variable.

"The FT teachers associated with Far West, Arizona, Bank Street, High/Scope, and EDC all report using (flexible/open classroom) practice to a significantly greater extent than the NFT comparison groups. In general, all Sponsors with the exception of Oregon and Pittsburgh have teachers who employ this practice to a greater extent than the NFT groups. This is an anticipated pattern for both of these Sponsors.

"...Arizona, Bank Street, Kansas, High/Scope, and SEDL have FT teachers who practice individualization of instruction to a significantly greater extent than their NFT comparisons. Only Oregon and Pittsburgh are below their respective NFT groups on this dimension." (p. II-34)

"...there are considerable Sponsor differences on the dimension of teacher-centeredness. Far West, Bank Street, and High/Scope have FT teachers who are significantly below their NFT comparisons on this dimension. Oregon and Kansas, on the other hand, have FT teachers who are significantly above their NFT groups. FT teachers in Pittsburgh and SEDL also report some degree of teacher-centeredness." (p. II-34)

"...Variability among the Sponsor groups (for reported teacher practices) is (considerable) and, in most cases, consistent with Sponsor models. For example, Far West, Arizona, Bank Street and High/Scope show profiles which are consistent with their programs which emphasize flexible/open classrooms, differentiation in terms of simultaneous instructional activities, and lack of teacher-centeredness. Variations in the size as well as the direction of the contrasts are also relatively consistent with each Sponsor's approach. The Kansas profile indicates that the individualized instruction is at the teacher's direction. Further, the reward system with its component of free play during part of the day is depicted as a comparatively open classroom by these FT teachers. Oregon is shown as a structured classroom, involved in large group work, where the teacher initiates most of the learning and playing activities. Pittsburgh shows a similar pattern, with more emphasis on adult supervision and choice of learning materials...EDC is shown as an open classroom with individualized instruction in comparison to its NFT group. Florida shows similarities to Sponsor 11. The SEDL teachers indicate a comparatively high degree of individualized instruction which is teacher initiated and supervised." (p. II-38)

Parent Involvement

"All FT teachers in all Sponsor groups placed greater importance on parent involvement in social and educational activities than did their respective NFT's.

"In general, all FT teachers value parent involvement in decision-making to a higher degree than their comparison groups.

"There is only one significant contrast with respect to teachers' reported use of parents for housekeeping duties, coming from Oregon FT teachers 'who indicate they do not use parents for housekeeping activities.' Only Kansas, EDC and SEDL appear to value parent involvement in such activities to any degree.

"...Teachers' opinions on the benefits that they, the children, or the program have received from parent involvement [vary markedly]. There are five significant Sponsor contrasts in which the FT teachers indicate more benefit from parent involvement than the NFT comparisons. These are Far West, Bank Street, Kansas, Florida and SEDL. Only Arizona, Oregon, and High/Scope have NFT teachers who identify a greater benefit from parent involvement than the FT groups." (p. II-41)

"...Sponsors' FT teachers express opinions which indicate a greater support for various aspects of parent involvement than shown by NFT groups. Both High/Scope and Pittsburgh have NFT comparison groups which involve parents in decision-making activities to a very small extent compared with other NFT groups. This tends to elevate those contrasts above those of Florida, a model which is based on parent involvement. The Florida profile suggests that the FT teachers using this approach make a relatively sharp distinction between involving parents in decision-making and involving parents with the pupil's learning process at home. In general, FT appears to be involving parents in decision-making to some degree. Further most FT groups identify more benefit from parent involvement than the NFT groups.

"It appears that Bank Street has the most active PAC's and Pittsburgh has the least active PAC groups. Florida, a parent educational model, is located near the mean of all Sponsors. It does not appear that there is a direct correlation between the amount of PAC activity and the importance teachers place on involving parents in the decision-making process." (p. II-41)

Teacher Satisfaction with Pupils' Progress

Neither the "cluster variable, teacher satisfaction with pupils' affective development [nor the] items concerning the teachers' satisfaction with progress in math [showed any overall significant differences between FT and NFT]. Only [the] variable teacher satisfaction with progress in reading produced a significant overall F test....three Sponsors' FT teachers report relatively more satisfaction with their pupils' [reading]: Arizona, High/Scope, and Pittsburgh. These three Sponsors have FT teachers who are also generally satisfied with the model. The FT teachers associated with Far West and Florida, on the other hand, are not as satisfied with their pupils' reading progress as the NFT teachers are although Florida teachers report satisfaction with the model in general." (p. II-46)

VI. SCHOOL STUDY (I.E., COMPARISON OF FT AND NFT SCHOOL CHARACTERISTICS) (Vol. II B, pp. III-1 - III-61)

A. Research Questions:

1. Overall, are FT and NFT schools similar?

2. Within sponsors, are FT and NFT schools similar?

3. Are there inter-sponsor, school and/or district differences?

B. Sample: 229 FT schools and 207 NFT schools were selected from among those schools at which children were tested during the 1971-72 school year. These were distributed across the ten sponsors involved in the pupil outcome analysis as follows:

Sponsor	No. of FT Schools	No. of NFT Schools
Far West	29	27
Arizona	34	37
Bank Street	32	25
Oregon	25	28
Kansas	23	18
High/Scope	14	13
Florida	32	20
EDC	17	17
Pittsburgh	12	14
SEDL	11	8
Total	229	207

"The number of schools available for description varies with the variable being studied because complete information was not available for each school." (p. III-8)

C. Variables: FT and NFT school differences (overall and within sponsor) are noted for the following variables:

1. School size and structure:

    a. entering grade in school

    b. highest grade in school

    c. total enrollment

    d. pupil-teacher ratios

2. School funds:
    a. whether Title I, other Federal, or state funds are received
    b. amount of Title I dollars per pupil
    c. amount of other funds per pupil
3. School services:
    a. number of medical professionals per hundred children
    b. number of social service professionals per hundred children
    c. number of psychological service professionals per hundred children
    d. number of instructional professionals per hundred children
    e. number of other professional specialists per hundred children
    f. number of non-professional personnel per hundred children
4. School district characteristics:*
    a. occupational structure
    b. percentage female in the labor force
    c. average family income
    d. average years of education
    e. percentage of persons who always lived in the same house
    f. percentage of persons below the poverty level
    g. percentage of families with children under 18 with female heads of households
    h. percentage Spanish-speaking and percentage non-English speaking persons

---

*FT/NFT differences are not calculated for these characteristics.

D. Analytical Method: Frequency distributions for overall FT/NFT contrasts were calculated. "In most cases, the frequency distributions for the total population were divided into five equal parts with approximately 20 per cent of the population in each category." (p. III-1) Histograms contrasting the percentage of FT and NFT units falling into each approximate quintile are presented. Means and standard deviations are reported for many intra- and inter-sponsor FT/NFT comparisons. District characteristics are considered in terms of inter-sponsor differences. No testing of differences is undertaken.

E. Comments: The major purpose of the analysis is to determine whether treatment groups are similar to comparison groups. Unfortunately, all overall FT/NFT, FT/NFT within sponsor, and inter-sponsor comparisons are drawn using either the district or the school as the unit of comparison. Observations concerning the quality of FT/NFT matching for child level analysis are inappropriate. The report concludes "The impact of these differences on pupil outcomes cannot be assessed until further studies merge school/community and pupil data." (p. III-61)

F. Results: Particular comparisons are reported in the table below. The report notes two special points.

"Two findings in this study are worthy of special note. First, we found little difference between FT and NFT in the proportion of schools receiving Title I and/or Federal or State funds. This finding suggests first of all that a reasonably close match was attained between the schools served by FT and NFT. In both cases, the schools serve low-income children eligible for Federal funding. This finding also poses some serious problems for the interpretation of FT/NFT contrasts. FT is not being compared to a "regular" school program, but rather to supplementary programs over and above the basic level of effort provided by the local educational agency. The quality of such programs varies widely, suggesting that additional information on the comparison programs would greatly enhance our understanding of pupil results.

"The second finding of note is the FT/NFT difference in the method of delivery of support services. The most available professionals in the FT world are the curriculum specialists, community liaisons, and social workers, whereas in the NFT world, specialist teachers are more common. In the FT model of delivery, the teacher or parent provides services directly to the children. The teacher, in turn, receives support services from the curriculum specialist and the parent from a social worker and/or a community liaison person. Hence, FT support services do not alter the classroom environment, but rather aid parents and teachers who are in direct contact with the children. The NFT model is more dependent on specialists and/or activities that service the children directly (i.e., supplementing or replacing regular classroom teaching).

"The FT method is designed to be less disruptive to the pupil-teacher relationship than the NFT. Furthermore, in the FT pattern of delivery, those children who may be in need of special services are not obviously segregated from the rest of the class. The FT method also extends beyond the school to the home. These two findings, as well as other differences in the schools and communities in which FT Sponsors operate, must be considered in interpreting pupil results." (pp. III-60 - III-61)

Summary of School Characteristics

Variable of Comparison	Overall FT/NFT Difference	Within Sponsor FT/NFT Differences	Inter-Sponsor Differences
School size & structure			
entering grade	no	a	a
highest grade in school	no	a	a
total enrollment	no	a	a
pupil/teacher ratio	FT < NFT	a	a
School funds			
receive Title I, other Fed. or state funds	no	a	a
Title I $/pupil	FT > NFT	FT > NFT	yes
Other $/pupil	FT > NFT	FT > NFT	yes
School services			
overall	FT > NFT	FT > NFT	yes
medical	FT > NFT	FT > NFT	a
social	FT > NFT	FT > NFT	a
psychological	FT > NFT	FT > NFT	a
instructional	FT > NFT	FT > NFT	a
other professional	FT < NFT	a	a
non-professional	FT < NFT	a	a
School District Characteristics			
occupational structure	b	b	yes
% female in labor force	b	b	no
average family income	b	b	no
average years of education	b	b	yes
% persons in same house	b	b	some
% persons below poverty level	b	b	yes
% female head of household	b	b	some
% Spanish speaking	b	b	yes
% non-English speaking	b	b	yes

[a] Comparison not commented upon in report.

[b] FT/NFT comparisons could not be drawn for district characteristics.

VII. FOLLOWING THROUGH ON HEAD START VARIATIONS: STUDY I (Vol. II B, pp. IV-1 - IV-24)

A. Research Question: Does continuity of curriculum approach from Head Start through kindergarten and first grade improve the school performance of participant children? (p. IV-14)

B. Sample: FT children in seven CIII sites which had previously been Head Start Planned Variation Sites. Traced children in these sites were divided into groups representing different patterns of preschool and early elementary experience. By site and pattern of experience, the numbers of children in the sample are:

	Duluth 0204	Tacoma 0213	Lincoln 0316	Portage-ville 0804	Ft. Walton 0902	Greeley 0906	Seattl 0910
HSPV-FT	41	11	28	34	26	17	12
Home-FT	42	10	110	20	16	9	23
HS-NFT (some unspecified HS)	17	2	9	23	4	12	2
Home-NFT	52	12	83	17	48	6	32

C. Variables:

1. Independent: four patterns of preschool and school experience:

   a. HSPV-FT

   b. Home-FT

   c. HS-NFT (some unspecified HS experience)

   d. Home-NFT

2. Dependent: Reading and math subtests for each of the following:

   a. Fall kindergarten WRAT

   b. Spring kindergarten WRAT

   c. Spring 1st grade MAT

D. Analytical Method: The outcome measures used in this study are subtest scores standardized relative to the entire NFT sample at each time point. Two approaches are used in this study. The first is simply a cross tabulation by sponsor, site, and pattern of preschool experience of income, mother's education, and outcome measures. The second is simply a presentation of the longitudinal patterns reading and math outcome scores which are then summarized by type (whether gain, even or loss for K and first).

E. Comments: Curiously, this study makes no use of HSPV pre- and post-test data. The criteria for classifying longitudinal patterns into gain-even-loss patterns are not made explicit. Two of the longitudinal patterns (for reading HSPV-FT Seattle, and math Home-FT Greeley) are unaccounatably left out of the summary. (p. IV-24) The study also does not make completely clear which reading and math scores are employed nor does it provide complete information on how these scores were standardized. (p. IV-14)

F. Results: "The most reasonable summary of the analyses described in the remainder of this chapter is that continuous HSPV-FT experience is not associated with academic performance at the end of kindergarten or at the end of first grade. These results are far from conclusive, however, and the character of the data forces us to withold judgement on the efficacy of the continuous experience when examined with these data." (p. IV-14)

"A few patterns deserve mention here:

--"Group mean income varies rather more within sites (among treatment groups) than among sites; some of this variability stems, no doubt, from the small sizes of some groups.

--"In five of the seven sites, the HSPV-FT group has the lowest annual family income of the four treatment groups. In five sites, moreover, the Home-NFT group has the highest family income.

--"Each site seems to have its own fairly characteristic mean levels of mothers' education, with Greeley, Fort Walton, and Portageville reporting generally less-school mothers than the other sites. The within-site ranks of the four treatment groups on this background variable follow no obvious pattern.

--"On all the test scores, moreover, we note substantial variability among sites and among treatment groups within sites." (p. IV-16)

"These displays provide little evidence that HSPV experience stimulated superior progress in kindergarten and first grade. Such evidence as we have suggests that any academic gains appeared early and tended to dissipate. The performance of HSPV children in kindergarten and first grade seems indistinguishable, on the basis of evidence at hand, from that of children who stayed home before kindergarten. While some kindergarten gains relative to NFT emerge, no HSPV-FT or Home-FT group gained in the first grade. The bulk of the changes were either non-significant or negative. The data give us no firm basis for discerning any difference between the performance of HSPV and 'home' as preschool treatments." (p. IV-16) (See also HSPV-FT Study II.)

VIII. FOLLOWING THROUGH ON HEAD START VARIATIONS: STUDY II
(Vol. II B, pp. IV-25 - IV-43)

A. Research Questions: How did the various sponsors, under the distinctive conditions of their various sites, perform relative to each other? (p. IV-25)

B. Sample: 137 CIII children in seven sites (four sponsors) who have been traced from HSPV through first grade of FT.

Site	0204	0213	0316	0804	0902	0906	0910
N	17	15	13	39	26	13	14

C. Variables:

1. Independent: sites.

2. Dependent:

   a. Seven HSPV pretest scores: Book 3D; Book 4A numbers and alphabet; PSI; Social responsiveness; associate vocabulary; numerical; and, sensory.

   b. Seven FT FK tests: PSI responsiveness/vocabulary, numerical concepts and sensory concepts; PPVT; WRAT, math, reading and spelling.

   c. Seven FT SK tests: Locus of control positive and negative; Gumpgookies, PPVT, WRAT math, reading and spelling.

   d. Five scores from FT S 1: MAT word knowledge, reading, math, "total reading", and MAT total.

D. Analytical Method: Site means on each of the outcome measures were ranked 7 through 1 (relative to the sites in this study only) for each outcome measure. Average site ranks for all "academic" test scores were calculated for each site for each time point (i.e., FH, FK, SK, S1).

Longitudinal patterns of site average ranks on academic tests were then graphed. Longitudinal patterns of site ranks for specific types of tests (reading, math, vocabulary) were also graphed.

E. Comments: This is an intriguing study since it is the first in the national evaluation to empirically link HSPV and FT data. Why the study omits HSPV spring test data is not clear, however. Given difficulties in interpreting such a ranking analysis, it is unfortunate that this study could not have been linked more directly to FT/NFT effects comparisons.

The study notes: "These findings obviously cannot support prescriptive conclusions, especially since they embody only comparisons among Sponsors' treatment groups, with no reference to "untreated" NFT or control groups. The FT/NFT contrasts reported elsewhere in this Follow Through report present rather a different picture of Portageville's accomplishments (this site showed clearcut gains in the ranking analysis), for example, compared to its own local controls. Portageville lost considerable ground during first grade in its reading scores, suggesting that some site-specific factor may have enhanced both FT and NFT reading progress in Portageville, making them both look good compared to FT and NFT in other sites but unimpressive compared to each other." (pp. IV-40 - IV-41)

F. Results: "...each Sponsor seems to have achieved, by the end of first grade, a roughly balanced set of gains with respect to the three principal dimensions of academic achievement. Putting it negatively, no Sponsor seems to be a "reading Sponsor" who neglects math, nor a "vocabulary Sponsor" who leans on rote memory achievements to the exclusion of more conceptual development. Each Sponsor's position in the post-test rankings of overall achievement and of each measured dimension of achievement appears to depend primarily on generalized entry level, and on site-specific factors, as would not be the case if Sponsors 'overgeneralized'." (p. IV-38)

"As we have noted here and there throughout this monograph [including both HSPV-FT studies I & II], we can really conclude from these data only that we wish heartily that certain additional bits of data had survived to the point of analysis. In the absence of usable data on non-Planned Variation (NPV) children in Cohort III, our analyses can neither support nor challenge the hopeful expectation that preschool sponsorship should confer long-term academic benefits on participating children. Nor can we declare some Sponsor's effects long-lasting and others' mere flashes in the pan. In brief, we have no real evidence on the effects of extended sponsorship in Cohort III. Some of the patterns that emerge from our crippled manipulations take suggestive forms, but the most interesting of them do not hold up in the context of the better-controlled Follow Through findings in the same sites." (p. IV-43).

1.4 Jane A. Stallings, David H. Kaskowitz, and others. Follow Through Classroom Observation Evaluation 1972-73. Menlo Park, CA: Stanford Research Institute, August, 1974.

This report is Stanford Research Institute's (SRI) final report on FT classroom observation evaluation.* It marks the culmination of some four years of effort to determine answers to a set of process questions:

> "(1) Are the observed classroom programs consistent with their sponsors' stated intentions? That is, does the model show a relatively high frequency of occurrence of those elements of the program that the sponsor rated as important?
>
> (2) Are the sponsored classrooms consistent within a site and between sites, i.e., do the four third-grade classrooms at Site A score similarly on specific program components, and do they resemble the third-grade classrooms at Sites B, C, D, and E?
>
> (3) Do sponsored classrooms differ on specified dimensions from the Non-Follow Through comparison classrooms?
>
> (4) How do Follow Through models differ from each other? How do they overlap, i.e., are there seven separate programs? How are they distinct from one another?" (p. 6).

The first two questions are examined by analyses summarized in part I (p. 196). Efforts to answer the latter two questions are summarized in II (p. 207).

The report also includes the results of efforts to obtain answers to three outcome questions:

> "(1) How do selected classroom processes relate to scores on the following: Metropolitan Achievement Test (MAT) (Reading and Arithmetic), Raven's Coloured Progressive Matrices, Coopersmith Self-Esteem Inventory, and Intellectual Achievement Responsibility Scale (IAR)?
>
> (2) How do selected classroom processes relate to such desired child behaviors as independence, cooperation, question-asking?

---

*The principal interim report was: Jane A. Stallings, Follow Through Program Classroom Observation Evaluation 1971-72 (Menlo Park, CA: Stanford Research Institute, August, 1973). Other reports and/or drafts had appeared earlier, for example, SRI, Follow Through Classroom Observations Draft (Menlo Park, CA: SRI, August, 1972). See also, Jane Stallings, Implementation and Child Effects of Teaching Practices in Follow Through Classrooms. Monographs of the Society for Research in Child Development, Serial No. 163, Nos. 7-8 (December, 1975), 133 pp.

(3) How do selected classroom processes relate to a child's school-absence rate?" (p. 7).

These outcome questions are addressed by analyses summarized in two sections of this summary:

III. INSTRUCTIONAL PROCESSES AS RELATED TO CHILD OUTCOMES--PARTIAL CORRELATIONS . . . . p. 221

IV. INSTRUCTIONAL PROCESSES AS RELATED TO CHILD OUTCOMES--MULTIPLE REGRESSION . . . . p. 237

A fifth section in this summary, THE COOLEY-LOWNES MODEL: COMMONALITY ANALYSIS, describes efforts to determine whether components in the instructional process are related to "reading and word analysis" outcomes at the end of the first grade (p. 246).

The analyses described in this summary rely upon measurements recorded with a Classroom Observation Instrument (COI). This instrument, prepared by SRI specifically for FT evaluation, consisted of three major sections. Since the COI and its constituent parts will be frequently referred to, we provide the following description taken directly from the report.

"1. Classroom Summary Information (CSI). The Classroom Summary Information section is designed to allow for the coding of the number of children enrolled and the number present on the observation day; (2) the number of teachers and aides assigned to the classroom; and (3) the number of volunteers or visitors present on the observation day. Also recorded in this section is the length of the school day.

"2. Physical Environment Information (PEI). The Physical Environment Information section provides space for coding information about the classroom setting, which includes the presence and use of specific equipment, instructional materials, games, toys, and displays in the classroom. The observer also records whether the classroom has movable tables and chairs and/or stationary desks in rows, whether seating is assigned during part of the day and/or or self-selected, and whether children are assigned to groups by the teacher or aide and/or select their own work groups.

"3. Classroom Observation Procedure (COP). The COP consists of three sections of codes that describe the classroom structure and process:

-- Classroom Check List (CCL)
-- Five-Minute Observation Preamble (PRE)
-- Five-Minute Observation (FMO)

The classroom observer completes the entire COP--the CCL, the PRE, and the FMO--approximately four times an hour or 16 to 20 complete COPs over a 4-5 hour observation day.

"a. Classroom Check List (CCL). As noted above, the Classroom Check List (CCL) is completed approximately four times an hour. The CCL is referred to as a 'snapshot' of the classroom because it allows for coding relatively stable 'pictures,' at a given point in time, of the activities engaged in and the distribution of the adults and children in the classroom. A record is made of activities occurring, of child and adult grouping patterns, of adult roles, of child involvement, and of the materials used in the academic activities (see Figure 1). For example, several activities may be taking place simultaneously in the classroom: a small group of children may be doing crafts under the guidance of a teacher aide, the teacher may be instructing a large group (over 8 children) in reading, three children may be working independently on a science project, and four children may be using trucks and blocks in independent play. In this case, all of these activities and the individuals involved in each would be recorded on the CCL.

"b. The Five-Minute Observation Preamble (PRE). The FMO preamble is designed to record information about the activity that the preselected focus person--whether child or adult--is actually engaged in (see Figure 2). This is done four times an hour. The observer focuses on the selected person just before the five consecutive minutes of observation begins and records: whether the focus person is an adult or child and his/her identification number; the CCL code number for the activity engaged in; the role of the adult--that is, whether the teacher, aide, or volunteer is observing, participating, or directing the activity, or not involved; the number of children involved; whether the activity in progress is a continuation of the previously observed activity (i.e., the preceding FMO); and the time the FMO was started. If, during the 5-minute observation period, there is a change in the activity in which the focus person is engaged, the observer records the activity actually taking place at the end of the time period in the designated space at the end of the FMO.

"c. The Five-Minute Observation (FMO). The last part of the COP, the Five-Minute Observation (FMO), is coded four times an hour immediately after the Preamble and is used to record interactions among people in the classroom. One person, the preselected focus adult or child, is followed throughout the 5-minute observation period and the activities and interactions of which he/she is a part, and only these, are recorded" (pp. 23-25).

I. ASSESSMENT OF IMPLEMENTATION (pp. 95-198)

A. Research Questions: "Two questions were addressed in this analysis:

-- Are the classrooms of a given model similar to each other on selected variables?

-- Are a sponsor's classrooms different from traditional classrooms on the implementation variables?" (p. 100).

B. Sample: "The primary considerations for the selection of sites were:

-- That they were among the sites where pupil testing was to occur in Spring 1973 as part of the Follow Through evaluation;

-- That each sponsor would then have a more balanced geographic distribution of sites, which included urban-rural and north-south projects;

-- That each sponsor would have at least two sites which he considered well implemented.

An attempt was made to select sites that previously had been observed so as to extend the longitudinal data base for possible future observation studies.

"Seven sponsors were selected for the study because they each had at least five sites that met the criteria. Six sponsors had five sites included in the study sample and University of Arizona had six sites, making a total sample of 36 sites. Within each Follow Through site, a sample of up to four first grade and up to four third grade classrooms were selected for observation, making a total of 136 first grade and 135 (Table 2) third grade classrooms observed. In each classroom, four children as well as the adults were observed. Table 2 lists the selected sites by sponsor and the number of classrooms observed by grade level. In addition, one Non-Follow Through classroom each at the first and third grade levels in each of the 36 sites (a total of 72 classrooms) were scheduled to be observed as comparison classrooms" (p. 11).

C. Variables: "From the several sections of the Classroom Observation Instrument, a list of variables intended to describe representative classroom elements was constructed by SRI staff for each sponsor's model. A Sponsor Variable Questionnaire was then compiled and sent to the individual sponsors for corrections, deletions, and additions" (p. 97).

Sponsors were asked "to rate the FMO variables as to (1) importance to the model and (2) expected frequency of occurrence relative to 'conventional' classrooms, according to the following definitions:

Table 2 (from Stallings, p. 12)

CLASSROOM OBSERVATION SAMPLE, SPRING 1973

Sponsor and Sites		Number of Follow Through Classes Observed	
		First Grade	Third Grade
Far West Laboratory for Educational R&D			
0201	Berkeley, Calif.*	4	4
0204	Duluth, Minn.*	4	4
0207	Lebanon, N.H.	4	4
0209	Salt Lake City, Utah	4	4
0213	Tacoma, Wash.	4	4
University of Arizona			
0305	Des Moines, Iowa	4	4
0307	Fort Worth, Texas*	4	4
0308	LaFayette, Ga.	3	4
0309	Lakewood, N.J.	4	4
0311	Newark, N.J.	4	4
0316	Lincoln, Nebraska	4	4
Bank Street College			
0502	Brattleboro, Vermont	3	3
0504	Fall River, Mass.	4	4
0506	New York City, P.S. 243K	4	4
0508	Philadelphia II, Pa.*	4	4
0510	Tuskegee, Ala.*	4	4
University of Oregon			
0703	E. St. Louis, Ill.	4	4
0707	New York City, P.S. 137K	3	3
0708	Racine, Wisc.	4	4
0711	Tupelo, Miss.*	4	4
0719	Providence, R.I.	4	4
University of Kansas			
0801	New York City, P.S. 77X*	2	2
0803	Philadelphia VI, Pa.*	4	4
0804	Portageville, Mo.*	4	3
0806	Kansas City, Mo.	4	4
0807	Louisville, Ky.	4	4
High Scope Educational Research Foundation			
0901	Greenwood, Miss.*	4	4
0902	Ft. Walton Beach, Fla.*	4	4
0903	New York City, P.S. 92M	3	4
0906	Greeley, Colo.	3	3
0907	Denver, Colo.	4	4
Education Development Center			
1101	Burlington, Vermont	4	4
1103	Philadelphia IV, Pa.*	4	4
1106	Paterson, N.J.*	4	4
1107	Rosebud, Texas	3	3
1108	Smithfield, N.C.	4	2
	Total	136	135

---

*These sites have been observed previously.

"(1) Importance to the Model--Rating Categories

Critical--This is an essential program component. Its absence in one of our classrooms would mean that the program was not implemented. It is emphasized in training the teaching staff.

Important--This program component is considered beneficial to children and the teaching staff is strongly encouraged to include this in the classroom; however, absence of this element would not indicate poor implementation.

Not Relevant--This is not a part of the curriculum we have developed. It is something with which we do not concern ourselves; the occurrence or lack of occurrence would not be important.

"(2) Expected Frequency Relative to Conventional Classrooms--Rating Categories

More--This should occur more frequently in our classrooms than in conventional classrooms.

Same--This should occur about as often in conventional classrooms as in our classrooms.

Less--This should occur less frequently in our classrooms than in conventional classrooms.

"Only variables that were considered critical and that should occur more often in the sponsors' classrooms were used as implementation variables. Sponsors were also asked to rank-order variables on the CCL. The amount of time allotted to each activity over a week's time was ranked, as were the materials to be used in the activities and the preferred groupings of children and adults. Finally the sponsors were asked to add observation variables which they thought might provide more explicit information regarding the implementation of their model. The variables that were finally selected for each model from the Sponsor Variable Questionnaire are used in the sections that follow to assess model implementation" (pp. 97-98).

Using the appropriate set of variables for each sponsor:

"Implementation scores for each sponsor were determined by rank-ordering the Non-Follow Through classroom mean scores on each sponsor variable and dividing the distribution into five equal parts. There are 35 Non-Follow Through first grades. The seventh lowest score is the first quintile cutpoint; the fourteenth lowest score is the second quintile cutpoint; the twenty-first lowest score is the third quintile cutpoint; and the twenty-eighth lowest score is the fourth quintile cutpoint. Any Follow

Through classrooms which have a score equal to or below the seventh score are in the first quintile. Any Follow Through classroom having a score above the twenty-eighth Non-Follow Through score is in the fifth quintile" (p. 103).

"Each sponsor's classrooms were assigned an implementation score according to the placement of the sponsor classroom means among the quintile cutpoints. If the classroom mean was greater than the fourth quintile cutpoint, an implementation score of 5 was assigned to the classroom; if the mean was less than or equal to the fourth quintile cutpoint, but greater than the third quintile cutpoint, a score of 4 was assigned. The quintile scores of 3, 2, and 1 were assigned in the same way. An implementation score of 5 indicates that the mean of the variable for the Follow Through classroom exceeded that of at least 28 classrooms, or 80%, or the Non-Follow Through classrooms. If a sponsor's classroom received an implementation score of 4 or 5, we could say that the classroom was on the upper end of the Non-Follow Through distribution; a score of 3 indicates that the classroom is in the mid-range of the Non-Follow Through distribution; and a score of 1 or 2 indicates that the classroom is on the lower end of the Non-Follow Through distribution" (p. 103).

"A total implementation score for a classroom was computed by summing the implementation scores across the corresponding sponsor's implementation variables and then dividing by the highest possible sum. The resulting proportion was then multiplied by 100 so that the total implementation score is expressed in terms of a percentage of the total possible. For example, if a hypothetical sponsor's classroom were being rated on four variables, the highest possible sum of implementation scores for a classroom would be 4 x 5 = 20. If a classroom had implementation scores of 3, 3, 4, and 5 on the individual implementation variables, then the total implementation score for the classroom would be (15/20)100 = 75%. The reader needs to understand that there is no zero point when computing implementation scores. If all classrooms received the lowest implementation score of 1 on every single implementation variable, their overall implementation score would be 20%, not zero. If they received scores of 5, their overall implementation score would be 100%. Thus, the actual range is from 20 to 100 and the midpoint is a score of 60.

"In order to compare the sponsor programs with Non-Follow Through classrooms, a total implementation score was also computed for each Non-Follow Through classroom on each sponsor's set of implementation variables. The mean and standard deviation of the Non-Follow Through pooled classrooms are reported for each sponsor" (p. 105).

In the analysis of FT/NFT differences, total implementation scores for each site were calculated by taking the average of the classroom total implementation scores at each site. In turn, these site-level, total implementation scores were used to calculate a single total implementation score for each sponsor.

D. Analytical Method: All analyses described below were performed separately for sponsors' first and third grade classes.

1. In order to determine whether a given sponsor's classroom activities were significantly different from the activities in NFT classrooms, a sponsor's overall total implementation score (the mean of all the classroom total implementation scores) was compared with the overall total implementation score of the NFT classrooms. T values and corresponding probabilities are reported.

2. ANOVA was used ". . . to test whether the sites differ on total implementation scores" (p. 106, emphasis added).

3. Histograms were used to compare FT and NFT classroom total implementation score distributions.

   Also reported is a large amount of descriptive information concerning each site's region, metropolitan status, population, median family income, percent of families below the poverty level, percent of population non-white, percent of adults over 25 who finished high school, average percent with preschool experience, and average percent with a first language other than English. These descriptive statistics do not enter into the analyses. They are provided only as an informal context for considering the results of the t-tests and ANOVA.

E. Comments: Perhaps the most important thing to be noted about this assessment of implementation is that it attempts to define models in terms of operational statements instead of operationalizing models in terms of the theoretical approaches they purport to represent. The problem of model specification is general to all of the FT evaluation and Stallings has dealt with it more thoroughly than any other FT investigator, but still the problem impairs the usefulness of implementation evaluation.* Specifically, findings from such an investigation cannot be abstracted with any confidence to the level of statements about this or that sponsor's "model" having been successfully implemented.

In addition to this general observation, several specific points warrant mention.

---

*As Weikart and Banet have pointed out, ". . . it is not necessarily the methodology that is at fault. . . . What is needed may be a clearer conceptualization of models, outcomes, and intervening processes, as well as better educational treatments for the researcher to evaluate." "Model Design Problems in Follow Through," in Rivlin and Timpane, Planned Variation in Education (Washington, D.C.: The Brookings Institution, 1975), pp. 70-71.

1. Sample--The lack of continuity in classroom enrollments introduces important questions that need to be considered in interpreting results.

> "The percentage of children for whom baseline WRAT test scores are available is an approximate indicator of the proportion of children within the observed classrooms who had been enrolled in Follow Through since their entry into school as kindergartners or entering first graders. As Tables 3 and 4 indicate, the percent of children with baseline data for first grade varies from a high of 81% in Newark to a low of 29% in Lakewood. In the third grade the range varies from a high of 79% in Smithfield to a low of 18% in Greeley" (p. 22).

Insofar as the number of FT children and the extent of their FT experience influences what goes on in the classroom, the migration of students in and out of FT classrooms poses a problem both for analysis and interpretation. Should a third grade classroom with 18% of its children having had a "consistent" FT experience be included in the same sample as one in which 79% of the children have been with FT since they entered school? Are both representative of the FT experience? Is overall teacher, aide, volunteer, and child behavior similar? Does a child with three years FT experience in the 18% classroom behave like a child with three years FT experience who is in a classroom where 79% of the children have had "consistent" FT experience? Do sponsors differ systematically in terms of the consistency of FT experience in their respective classes? The answers to these questions are not clear cut, but the reader should keep them in mind when considering the results reported below.

2. Variables--The construction of the implementation variable involves an immense reduction of information on two accounts.

First, the 100,000 possible coded combinations of the Classroom Observation Instrument which describe classrooms, their contents, and activities are reduced to 602 constructed variables (p. 41 and Appendix F). From these 602, an undisclosed number are selected for "a list intended to describe representative classroom elements . . . for each sponsor's model" (p. 97).* These lists are modified according to sponsor responses as described earlier and a set of 55 variables is arrived at (see Table 11 above). The use of this set of 55 variables for implementation assessment implies that it is adequate to the task of describing all the important elements of each sponsor's approach. Even casual inspection of the list suggests that it is not adequate to describe what goes on in a first or third grade classroom. FT/NFT differences and within-sponsor consistency are consequently examined in terms of a very circumscribed set of variables.

---

*The lack of specific decision rules, with accompanying rationale, for the two stages in the reduction from 100,000 to an undisclosed but less than 602 variables is a matter of some concern itself. However, the extent of the reduction is of primary concern here--not the arbitrary composition of the list of variables sponsors were asked to rate.

Second, the total implementation score is the product of another reduction process. In this case, the reduction is a process of mathematical aggregation involving a number of steps. (1) Coded individual behaviors were summed to obtain a classroom score on a given variable. (2) Then (after FT classroom scores were recoded with reference to the distribution of NFT classroom scores), classroom scores on a set of variables were normalized to produce a single implementation score for each classroom. (3) The aggregation process is taken one step further when the implementation scores, one for each classroom, are averaged to generate a single implementation score for every site. The analysis of within-sponsor consistency of implementation, is performed on these site Total Implementation Scores (TIS). The aggregation process doesn't stop at this point. (4) It is carried one step further when the site TIS's are averaged to give a sponsor TIS. The sponsor TIS is used for the determination of FT/NFT differences.

It's difficult to say what the product of these many levels of aggregation really means. Consequently, it is difficult to interpret the results of the statistical analyses performed on the products of that aggregation.

Another point should be noted concerning the construction of the variables that were used. Each sponsor was scored on the basis of from 17 to 29 variables. The reduction of these separate scores to a single classroom score involved summing across a number of variables that are considerably different from one another. For example, calculating a class implementation score for University of Kansas involved adding scores for:

-- the percentage of time spent on numbers, math, and arithmetic,

-- the percentage of time during which an aide works with a small group,

-- the quantity of teacher, aide, or volunteer positive, corrective feedback to children,

and fourteen other variables, and dividing by the maximum score that could have been received on all of the FT variables. Reflection on the procedure leads one to ask: Is there any common conceptual dimension underlying these variables? If so, what is it? If not, what does the sum, or the average of these scores represent? Since the classroom implementation scores are the material from which site and sponsor TIS's are calculated, the same questions need to be asked concerning those two measures.

Another difficulty is encountered in interpreting the implementation measure. As noted earlier, FT classroom scores on respective variables were recoded onto a 1 to 5 scale based on the distribution of NFT classroom scores on the same variables. This procedure works fine when the 35 NFT classroom scores on a given variable are distributed in such a manner that they can be divided into quintiles. However, anytime more than 14 (20%) of the NFT classrooms have a zero score, the lowest quintile cannot be neatly defined. Similarly, if more than 21 NFT classrooms had a zero score on a given variable, the bottom two quintiles cannot be defined. SRI's report notes that in situations of this sort, ". . . the implementation score assigned to that classroom corresponds to the highest quintile cutpoint that was zero" (p. 104). The following table summarizes the procedure:*

No. of NFT Classrooms With a Zero Score	Value Assigned to a FT Classroom With a Zero Score
> 14	2
> 21	3
> 28	4
all 35	?

In effect, the decision rules used by Stallings defined four, three, and two point distributions for recoding instead of just the five point scale implied by dividing NFT scores into quintiles. Specifically, the following scales were defined:

5, 4
5, 4, 3
5, 4, 3, 2
5, 4, 3, 2, 1

---

*Among third grade classrooms, NFT scores on six of the 55 implementation variables were distributed in such a way that all four cutting points were zero. These variables were: 70, sewing, cooking, pounding; 74, practical skills acquisition; 450a, all child open-ended questions; 469a, all adult reinforcement with tokens; 514c, two children working together, using concrete objects; and 515c, small group working together, using concrete objects.

Among first grade classrooms, NFT scores on two of the 55 implementation variables were distributed in such a way that all four cutting points were zero. These variables were: 514a and 515c.

The report indicates in Appendix M, Table M-1, that NFT classroom scores on 19 variables (out of the 55 used) were distributed in such a way that unique (i.e., non-zero) cutting points between one or more quintiles could not be defined.* Consequently, the calculation of a mean involves summing scores on five, four, three, and two point scales but dividing the sum as though they were all on a five point scale.

Of no less significance than the points already touched upon is the issue of the stability of the quintile cutting points. The instability of scores is due inter alia to:

-- "the number of classrooms observed and the procedure of sampling classrooms," and

-- the "sampling of classdays for each Follow Through classroom" (pp. N-3 to N-4).

Stallings suggests that each cutting point which divides the distribution of NFT classroom scores into quintiles can be considered an estimate of a true cutting point. Confidence limits can be cast about these estimates. If the confidence intervals about two adjacent points are fairly small, then one can be confident that the quintile they demarcate is fairly well-defined. If the confidence intervals around all four cutting points are fairly narrow, then a relatively well-defined set of quintiles is obtained, one that would serve as a good basis for a five point scale.

Using estimates of variability due to the sampling of classrooms, SRI investigators calculate 95% confidence intervals for the quintile estimates expressed in percentiles of the classroom distribution. The results of this inquiry are summarized as follows:

Quintile Cutpoint	Corresponding Percentile	Confidence Interval
1	20	6-31
2	40	22-61
3	60	39-78
4	80	69-94

It's obvious that the cutting points are anything but stable. For example, if one were to recode a FT classroom score on a given variable on the basis of NFT classrooms' distribution on the same variable, and that score lay above the second cutting point but below the third cutting point on the NFT classroom distribution, it would be recoded as a 3 since it would fall within the third or middle quintile. However, the results just reported indicate that the lowest confidence limit of the second cutting point is equivalent to the 22nd percentile and the highest confidence limit of the third cutting point is equivalent to the 78th percentile. In other words, the confidence intervals demarcating the middle quintile spans nearly all the second and fourth as well as the third quintile. A recoded FT classroom score of three could be a two or a four with an uncomfortably high degree of probability. The same applies to all the cutting points, and the assignment of recoded FT values based on these very unstable estimates could be off as much as two points (except in the case of scores falling within the highest or lowest quintile).

---

*This was true on 20 variables for third grade.

The difficulty presented by the instability of the cutting point estimates due to the sampling of classrooms is compounded by the instability due to the sampling of class days. There was some day-to-day variability in the classroom observation scores. When this variability is taken into consideration, ". . . the estimated implementation score will be within one of the true score" (N-9). That is, FT classroom scores recoded to a five point scale based on the distribution of NFT classroom scores may be off as much as one point in either direction.

When both sources of instability are taken into consideration:

"The compound effect of estimating the Follow Through classroom value and estimating the Non-Follow Through quintiles is that the estimated implementation score assigned a particular Follow Through classroom may be different from the true score by as much as two points. In the worst situation, a classroom that has a true implementation score of 3 appears to have a substantial chance of getting an observed score on an individual variable of from 1 to 5" (p. N-9).

3. Analytical Method--The purpose of the analysis is to determine whether sponsors have been successful in implementing and maintaining classroom activities that are both consistent across a sponsor's sites and different from traditional classroom activities. As the study is designed, inferences about sponsors' success are hardly warranted. The reason is that the cross-sectional nature of the design precludes controlling for causes other than sponsors' efforts to make FT classrooms different from NFT classrooms. Similarly, the consistency of scores across a given sponsor's sites would be more safely interpreted as a descriptive finding.

   A second point to keep in mind when reviewing the results is the fact that analyses were conducted with data aggregated to either the site or sponsor level. Thus, although the research questions are formulated in terms of the classroom (p. 100) and the summary discusses intra-site classroom variability, findings at best warrant conclusions concerning either (a) inter-site consistency of implementation or (b) an intra-sponsor FT/NFT difference.

F. Results: Readers are referred to the report for details. We summarize here only the general findings (see next page).

On the basis of the measures and methods used, it would appear that FT classroom activities differ from those in NFT classrooms. The issue of consistency of implementation is another matter. "All sponsors except Far West Lab and University of Kansas and High/Scope have some classrooms which have scores below the Non-Follow Through mean" (p. 195). All sponsors, with the exception of Bank Street, manifest significant inter-site variability at either the first or third grade level.

SUMMARY OF IMPLEMENTATION ASSESSMENT RESULTS

	A Significant FT/NFT Difference[a]	Significant Differences Among Sites[b]
Far West Labs		
first grade	yes ($p < .001$)	NS
third grade	yes ($p < .001$)	yes ($p < .05$)
University of Arizona		
first grade	yes ($p < .001$)	yes ($p < .001$)
third grade	yes ($p < .001$)	yes ($p < .01$)
Bank Street College		
first grade	yes ($p < .001$)	NS
third grade	yes ($p < .001$)	NS
University of Oregon		
first grade	yes ($p < .001$)	yes ($p < .001$)
third grade	yes ($p < .001$)	NS
University of Kansas		
first grade	yes ($p < .001$)	yes ($p < .01$)
third grade	yes ($p < .001$)	NS
High/Scope		
first grade	yes ($p < .001$)	yes ($p < .001$)
third grade	yes ($p < .001$)	yes ($p < .001$)
EDC		
first grade	yes ($p < .001$)	yes ($p < .01$)
third grade	yes ($p < .001$)	yes ($p < .05$)

---

[a] Based on a t-test of the difference between the mean of all the classroom total implementation scores for a given sponsor and the mean of all the 35 NFT classroom total implementation scores.

[b] Based on (presumably) a one-way ANOVA to test whether the site implementation scores are equal.

Source: Stallings, pp. 111, 123, 137, 147, 159, 171, and 185 for each of the seven sponsors respectively.

II. THE NATURE OF DIFFERENCES AMONG SPONSORS (pp. 199-232)

A. Research Questions: A set of analyses addresses the following four questions:

"-- What, if any, are the major dimensions that differentiate the classrooms of the several sponsors?

-- What classroom observation variables contribute to these dimensions?

-- How are sponsors differentiated on these dimensions?

-- To what degree can classrooms be identified with their correct sponsor by using the classroom observation data?" (p. 199).

B. Sample: "Since only differences among sponsors were of interest, Non-Follow Through classrooms were not included in the discriminant analysis phase.

"Table 89

NUMBER OF CLASSROOMS INCLUDED IN THE DISCRIMINANT ANALYSIS BY SPONSOR AND GRADE LEVEL

	First Grade	Third Grade
Far West Laboratory	20	19
University of Arizona	22	24
Bank Street College	18	18
University of Oregon	19	18
University of Kansas	16	15
High/Scope	18	19
EDC	19	17
Total	132	130" (p. 202)

C. Variables:

1. Independent--Variables were selected from the Physical Environment Information (PEI), Classroom Check List (CCL), and Five-Minute Observation (FMO) portions of the Classroom Observation Instrument for the construction of six canonical variables. Forty-five variables taken from the PEI and CCL were classified as "environmental variables." "These environmental variables describe the grouping patterns of adults and children, the relative amount of time spent in certain activities, and the physical equipment used. The analysis of the (32) selected FMO variables permitted determination of the classroom interaction variables distinguishing the sponsors" (p. 201).

The variables were selected on the basis of three criteria:

"(1) Results from range tests of differences among sponsor means, performed for 1971-1972 analyses--To some extent evidence was already available on which variables differentiated sponsors. The discriminant analyses would add to this by providing information on the relative weights of these variables.

"(2) The implementation variables employed extensively in earlier sections of this report--Those variables which the sponsors are attempting to influence in the classroom may be very closely linked to the characteristics that differentiate sponsors.

"(3) The selection of four or five variables for each sponsor, for a total of between 30 and 35--Unless there were several variables advocated by each of the sponsors, some important differences were likely to be ignored" (p. 201).

Nineteen environmental variables were used to discriminate among sponsors in the first grade and seventeen in the third grade. Fifteen FMO variables were used to discriminate among sponsors in the first grade and fourteen in the third grade.

2. Dependent-- Classification according to association with respective sponsors.

D. Analytical Method:

"The analysis we perform here consists of two parts. In the first part, we perform a multiple discriminant analysis. (See Cooley and Lohnes, 1962, and Rao, 1965, for details concerning this technique.) From a set of classroom observation variables, new variables, called discriminant functions, are generated, each of which is a linear combination of the original COI variables.* These discriminant functions have the property that they maximize the variability among sponsors relative to within-sponsor variability. The discriminant functions can be viewed as major dimensions on which sponsors or groups of sponsors are distinguished.

"The second part of the analysis consists of a classification procedure. (See Anderson, 1958, for details concerning this technique.) The 'best' (most effective) linear functions for classifying classrooms by sponsor, under multivariate normal theory, are derived based on selected COI variables. Then, based on these classification functions, each classroom is matched with the sponsor it was most likely to have come from; i.e., we act as if we do not know which sponsor the classroom is affiliated with and we identify the sponsor the classroom is most likely to have come from. This operation determines how reliably a classroom can be identified with the correct sponsor or, conversely, how often it is confused with the program of another sponsor. Obviously, if the sponsor of a classroom can be identified correctly in this way, then differences among sponsors are quite apparent, and systematic observation by means of the COI can reveal them" (p. 200).

"The stepwise discriminant analysis program, BMD 07M, from the Biomedical Statistical Package of programs (BMD Biomedical Computer Program, 1973) was used. The classroom was the unit of analysis. In a preliminary phase, the program selects a subset of the variables for inclusion in each analysis. Variables are entered for inclusion one at a time until none of the variables that remain differentiate among sponsors" (pp. 201-202).

Four separate discriminant analyses were undertaken:

(1) Selected environmental variables, first grade;
(2) Selected environmental variables, third grade;
(3) Selected FMO variables, first grade; and
(4) Selected FMO variables, third grade.

E. Comments:

1. "Portions of the discriminant analysis and the classification analysis are based on assumptions of multivariate normality and homogeneous within-sponsor covariance matrices. . . . Many of the classroom observation variables have clearly nonnormal distributions. It is also doubtful that the within-sponsor covariance matrices are homogeneous. Hence, the optimal properties of these techniques are only roughly approximated, at best. In such circumstances, these analyses should be looked upon as heuristic exercises for trying to achieve a better understanding of the differences among sponsors" (p. 200, our emphasis).

2. The discriminant functions were able to differentiate sponsors lying at extreme ends of some scales. However, they usually left a bundle of three to five sponsors undifferentiated in the middle of the scale. In no case were discriminant functions able to differentiate all seven sponsors one from another.

3. For some unexplained reason, NFT classrooms were not included in the discriminant analyses. Consequently, an important opportunity to determine whether sponsor classrooms could be differentiated from NFT classrooms was lost.

4. At some point a decision was made to use the particular descriptions of all sections of the COI. It has never been made clear why the specific descriptions were chosen. Moreover, at some point in the evaluation process, a decision was made

to reduce the total number of observation measurements to a more managable set for use in discriminant analysis. Information on grounds for doing this is not provided. Hence, it is difficult to evaluate the rationale for selecting the 45 environmental variables and the 32 activity variables that were selected.

F. Results:

1. Major Dimensions That Differentiate Sponsors. Since in each of the four analyses ". . . the first three discriminant functions account for over 80% of the total discriminating power that can be achieved with all six possible functions," the report only talks about those three functions (p. 204). Results concerning differentiating dimensions are reported and summarized with reference to the four analyses performed.

   a. Environment Variables--First Grade.

   "The first discriminant function relates to the configuration of adults and children in the classroom" (p. 206). High positive and negative coefficients appear to identify a scale that runs ". . . from a less structured configuration where children tend to work independently and work with adults on a personalized basis (one or two children at a time) to a structured classroom configuration where children tend to work with adults in small or large groups" (p. 206). The "University of Kansas and the University of Oregon are at [the negative] extreme . . . and Far West and EDC are at the other extreme. . . ." (p. 206).

   Interpretation of the second discriminant function is beset by an "anomaly in the data" that appears to be due to mistaken coding (p. 208). This function serves mainly to separate the University of Oregon from the rest of the sponsors on some unidentified dimension.

   The third discriminant function is likewise difficult to interpret. Far West lies " . . . at the positive end of the continuum and the University of Arizona at the negative end" (p. 210). The conceptual dimension underlying the function is not identified.

   b. Environmental Variables--Third Grade.

   "The first discriminant function relates to the configuration of the classroom and the amount of arts and crafts" (p. 210). "The way sponsors were aligned along the first dimension was very similar to what was found for the first grade with a reversal in direction. University of Kansas and University of Oregon are at one extreme (with more small and large groups with aides in math) . . . and Far West and EDC are at the other extreme (with more arts and crafts and a wide variety of activities over the day). . . . (p. 210).

"The second function also relates to the configuration of the classrooms" (p. 210). "High/Scope, University of Arizona, and University of Kansas are on the positive end of the scale with University of Oregon and EDC on the negative end of the dimension" (pp. 210-211).

"The interpretation of . . . [the third discriminant function] . . . is unclear and it does not seem to have a counterpart in any of the three dimensions at first grade level. This dimension would cluster Bank Street, the University of Kansas, and EDC at the positive end of the continuum and the other four sponsors at the negative end" (p. 212).

c. FMO Variables--First Grade.

The first discriminant function differentiates University of Kansas classrooms from University of Oregon classrooms and these two sponsors from the other sponsors. "This discrimination is accomplished primarily by two items (adult reinforcement with token, academic and child group response to adult academic command, request, or direct question" (p. 212).

The substantive interpretation of the second function is likewise unclear. The second function appears to separate the University of Kansas and the University of Oregon from the other sponsors.

Just how the University of Arizona differs from the other sponsors is not clear, but the third discriminant function indicates that it is.

d. FMO Variables--Third Grade.

"The first discriminant function appears to differentiate sponsors on a dimension related to the degree of child interaction with adults and the child's role in the interactions" (p. 215). "University of Arizona and EDC are at the extreme negative end of the dimension . . . [and] . . . University of Kansas and University of Oregon are at the extreme positive end. . . ." (p. 218).

The dimension underlying the second function is not identified. The University of Oregon is differentiated from other classrooms in the positive direction with the University of Kansas differentiated in the negative direction.

> "Extensive use of questions and feedback by adults seems to be the most characteristic feature of the negative end of . . ." the dimension underlying the third function (p. 220). "The third FMO function at third grade level bears considerable similarity to the third FMO function at first grade level, except that the direction of the function is reversed (i.e., the signs of the weights are in opposite directions and the sponsor means have opposite signs)" (p. 218). "The University of Arizona is at the positive extreme of the third function, with Bank Street fairly close to it. Far West, High/Scope and EDC are clustered at the negative end of the third function" (p. 220).

The foregoing results are summarized in the table which follows this page. In summarizing these results, the report makes the following important point: "With the possible exception of a 'degree of structure' dimension, the discriminant functions on which the sponsors differed did not seem to represent abstract concepts, such as 'individualization of instruction' or 'child initiation' versus 'responsiveness' but were dominated by one or two very specific classroom process variables, such as 'Large group with aide/Math' or 'Adult reinforcement with token, academic.' One or two individual variables shared by one or two sponsors seem to be the best discriminators" (p. 228).

This point is especially important: Insofar as sponsors are differentiated by the analyses, the differentiation is largely due to one or two idiosyncratic aspects of a few sponsors' classrooms and lack thereof among other sponsors. In terms of the first research question, this implies that, with the exception of the dimension of structure, the discriminant analysis identifies no "major dimensions" of generalizable interest, none that would help interpret how sponsor classroom activities really differ.

2. "Observational Variables" that contribute to differentiating "major dimensions". Since "structure" was the only identified major dimension differentiating sponsors, the answer to the second research question is logically constrained. The report treats the matter in the following manner:*

> "The next question was: What observation variables contribute to these dimensions? This must be answered separately for the PEI/CCL analyses and FMO analyses. For the PEI/CCL analysis, the first and third grade results are very similar. Table 99 shows the observation variables with high scaled coefficients on the first two dimensions.

---

*The tables referred to appear on the following two pages.

SUMMARY OF RESULTS OF THE DISCRIMINANT FUNCTION ANALYSIS

	Dimension Upon Which Function Discriminates Sponsors	Differentiated Sponsor
I. Environmental Variables		
A. First Grade		
1. First function	"configuration of adults and children in the classroom--non-structured, high adult child ratio"	U. of Kansas and U. of Oregon at negative extreme; Far West and EDC at positive extreme.
2. Second function	unidentified	U. of Oregon different from rest.
3. Third function	unidentified	U. of Arizona at negative extreme; Far West at positive extreme.
B. Third Grade		
1. First function	"configuration of classroom and the amount of arts and crafts"	U. of Kansas and U. of Oregon differentiated from Far West and EDC.
2. Second function	also configuration in classroom, but difficult to specify	High/Scope, U. of Arizona, an U. of Kansas at positive end with U. of Oregon and EDC at negative end.
3. Third function	unidentified	Bank Street, U. of Kansas, an EDC at the positive end versu the other four at the negativ end.

SUMMARY OF RESULTS OF THE DISCRIMINANT FUNCTION ANALYSIS

	Dimension Upon Which Function Discriminates Sponsors	Differentiated Sponsor
II. FMO, Classroom Activity Variables		
A. First Grade		
1. First function	unidentified	U. of Kansas and U. of Oregon different from the rest.
2. Second function	unidentified	U. of Kansas and U. of Oregon different from the rest.
3. Third function	unidentified	U. of Arizona different from the rest.
B. Third Grade		
1. First function	"degree of child interaction with adults and the child's role in the interaction"	U. of Arizona and EDC at the extreme negative end versus U. of Kansas and U. of Oregon at the extreme positive.
2. Second function	unidentified	U. of Oregon at positive extreme, U. of Kansas at the negative extreme, both differentiated from others clustered in the middle.
3. Third function	"extensive use of questions and feedback by adults"	U. of Arizona and Bank Street at positive end with Far West, High/Scope and EDC at the negative end.

"The results for the FMO analyses (Table 100) differed between first and third grades. The fact that the University of Kansas classrooms phase out their use of material rewards (tokens) by the third grade may partially explain the differences. In first grade, the first canonical variable separates University of Kansas and University of Oregon from one another on the basis of tokens versus group responses. The second separates both sponsors from the others.

"In the third grade, the first discriminant function does the same thing (University of Kansas and University of Oregon versus others) as the second discriminant function does in first grade. The 'tokens' variable no longer distinguishes University of Kansas from University of Oregon, and another difference among sponsors dominates" (p. 228).

3. Manner in which sponsors are differentiated. "In all four analyses and in the classification procedure, the classrooms associated with the University of Oregon and the University of Kansas models tended to cluster apart from each other and apart from other sponsors. In a separate analysis without these two sponsors' classrooms, the University of Arizona was distinguished from the remaining four sponsors on the basis of Var. 454a (Child's extended response to questions) and Var. 440a (Adult communication or attention focus, small group). High/Scope was differentiated from the other four on the basis of variables indicating a high level of verbal interaction between adults and children and, as would be expected from their Piagetian model, Var. 510c (Child self-instruction, objects). The third discriminant function distinguished Far West Lab from EDC and Bank Street on the basis of child questioning and all adult praise" (p. 228).

4. Degree to which classrooms can be accurately identified using classroom observation data. Table 97 reports the results of the classification procedure. "Of the 132 first grade classrooms, 83% were correctly categorized on the basis of the nineteen CCL and PEI variables and 78% were correctly categorized on the basis of the fifteen FMO variables. Of the 130 third grade classrooms, 73% were correctly categorized on the basis of the seventeen CCL and PEI variables and 79% were correctly categorized on the basis of the fourteen FMO variables. The number of classrooms that were misclassified ranged from 23 to 34" (p. 222).

   The ". . . University of Kansas and University of Oregon programs are most distinct. Their classrooms are rarely identified with them. A high percentage of the classrooms of the University of Oregon and University of Kansas programs are correctly classified on every analysis performed. The rate of correct classification was at least 85% and reached

Table 99

ANALYSES OF PEI AND CCL VARIABLES

A. First discriminant function

	First Grade Variables	Third Grade Variables
Large positive weights	All children independent	All children independent
	Personalized instruction in reading	Teacher with one child
	Puzzles, games/academic activities	Arts, crafts
Large negative weights	Small group of children with any adult	Small group with aide/Math
	Large group with aide/Math	Large group with aide/Math

B. Second discriminant function

	First Grade Variables	Third Grade Variables
Large positive weights	Small group of children with any adult	Wide variety of activities, ove one day
	Large group with aide/Math Numbers, math, arithmetic Aide with small group	Small group of children with a adul
Large negative weights	Math or science equipment/Academic activities	Numbers, math, arithmetic
	Wide variety of activities, over one day	All children independent

Table 100

ANALYSES OF FMO VARIABLES

A. First discriminant function

	First Grade Variables	Third Grade Variables
Large positive weight	Adult reinforcement with token, academic	Child questions to adult
Large negative weight	Child group responses to adult academic command, request, or direct question	Child group responses to adult academic command, request, or direct question All adult corrective feedback to children

B. Second discriminant function

	First Grade Variables	Third Grade Variables
Large positive weight	Adult communication or attention focus, large group	Adult communication or attention focus, one child
Large negative weight	Child group responses to adult academic commands, request or direct question All adult praise to children	Child group responses to adult academic command, request, or direct question

Table 97

CLASSIFICATION OF CLASSROOMS BY SPONSOR

PEI and CCL Variables, First Grade

	Classrooms Identified as Sponsor						
Sponsor	Far West	Univ. of Arizona	Bank Street	Univ. of Oregon	Univ. of Kansas	High/ Scope	EDC
Far West	18	2	0	0	0	0	0
University of Arizona	0	17	2	0	0	2	1
Bank Street College	1	2	13	0	0	1	1
University of Oregon	0	0	0	19	0	0	0
University of Kansas	0	0	0	0	14	2	0
High/Scope	0	1	2	1	1	13	0
EDC	2	0	2	0	0	0	15
Non-Follow Through	6	6	5	5	0	2	10

PEI and CCL Variables, Third Grade

	Classrooms Identified as Sponsor						
Sponsor	Far West	Univ. of Arizona	Bank Street	Univ. of Oregon	Univ. of Kansas	High/ Scope	EDC
Far West	15	2	2	0	0	0	0
University of Arizona	0	17	2	0	0	5	0
Bank Street College	3	1	11	0	0	0	3
University of Oregon	0	0	0	17	1	0	0
University of Kansas	0	0	1	1	13	0	0
High/Scope	2	3	0	0	1	12	1
EDC	2	0	5	0	0	0	10
Non-Follow Through	10	0	7	16	0	0	3

FMO Variables, First Grade

	Classrooms Identified as Sponsor						
Sponsor	Far West	Univ. of Arizona	Bank Street	Univ. of Oregon	Univ. of Kansas	High/ Scope	EDC
Far West	15	1	3	0	0	0	1
University of Arizona	0	19	2	0	0	0	1
Bank Street College	2	2	10	0	0	1	3
University of Oregon	0	0	0	19	0	0	0
University of Kansas	0	0	0	0	16	0	0
High/Scope	2	0	0	1	0	13	2
EDC	3	1	1	0	0	3	11
Non-Follow Through	9	5	8	0	0	6	6

FMO Variables, Third Grade

	Classrooms Identified as Sponsor						
Sponsor	Far West	Univ. of Arizona	Bank Street	Univ. of Oregon	Univ. of Kansas	High/ Scope	EDC
Far West	13	3	0	0	0	3	0
University of Arizona	1	20	0	0	0	1	2
Bank Street College	0	1	15	0	0	1	1
University of Oregon	1	0	0	16	0	0	1
University of Kansas	0	0	1	0	14	0	0
High/Scope	0	3	1	1	0	9	5
EDC	0	0	0	0	0	1	16
Non-Follow Through	6	1	7	4	1	10	7

100% in three cases. Very few classrooms from other sponsors are misclassified as University of Oregon or University of Kansas classrooms" (p. 225).

The report addressed the issue of classification another way--by computing the number of misclassifications obtained. The results appear in Table 98.

"There were 114 instances of misclassification, or 22% of the 524 classifications made over the four analyses. . . . University of Oregon and University of Kansas classrooms were rarely misclassified and the classrooms of other programs were rarely classified in University of Oregon and University of Kansas programs. There are five instances of High/Scope misclassified as University of Oregon or University of Kansas (see High/Scope row of Table 98) but none of the other four sponsors are ever misclassified as University of Oregon or University of Kansas.

"The classification functions developed for the Follow Through classrooms were used to classify the Non-Follow Through classrooms in order to examine the relative degree to which the processes in Non-Follow Through classrooms are similar to the various Follow Through programs. The results are displayed in the last row entry for each data set in Table 97" (p. 227).

"The Non-Follow Through classrooms were most often classified as Far West, University of Arizona, Bank Street, University of Oregon, and EDC classrooms. On the third grade physical environment and checklist of classroom grouping and activities, the Non-Follow Through classrooms were classified as University of Oregon classrooms 16 out of 36 times. It is notable that the Non-Follow Through classrooms were rarely classified as either University of Kansas or High/Scope classrooms" (p. 231).

Table 98

NUMBER OF MISCLASSIFICATIONS AND CORRECT CLASSIFICATIONS SUMMED OVER THE

	Classrooms Identified with the Wrong Sponsor						
Sponsor	Far West	Univ. of Arizona	Bank Street	High/ Scope	EDC	Univ. of Oregon	Univ. of Kansas
Far West	X	8	5	3	1	-	-
University of Arizona	1	X	6	8	4	-	-
Bank Street College	6	6	X	3	8	-	-
High/Scope	4	7	3	X	8	3	2
EDC	7	1	8	4	X	-	-
University of Oregon	1	-	-	-	1	X	1
University of Kansas	-	-	2	2	-	1	X
Total							

---

[*] Each classroom was classified twice--once using the PEI-CCL variables and once ables--for a total of 524 classifications in first and third grades.

p. 226

FOUR DATA SETS*

Total Wrongly Classified	Total Correctly Classified
17	61
19	73
23	49
27	47
20	52
3	71
5	57
114	410

using the FMO vari-

III. INSTRUCTIONAL PROCESSES AS RELATED TO CHILD OUTCOMES--PARTIAL CORRELATIONS (pp. 233-326)

The relationship between instructional process variables and selected child outcome measures was analyzed three different ways. This summary deals with the first technique used, first order partial correlations. The following two summaries deal with other techniques.

A. Research Question: Are selected child outcome measures--both achievement and behavior--significantly related to instructional process variables when the influence of entry WRAT scores is partialled out?

B. Sample: One hundred and sixty-six classrooms were selected on the basis of the criterion that "at least 20% of the children in the classroom had to have baseline test data" (p. 233). References in the text indicate that for the most part first graders included in this analysis were Cohort III, entering kindergarten children. Third graders were for the most part Cohort I, entering kindergarten students. The distribution of classrooms across sites and sponsors is indicated in Table 101 of the report (p. 234).

C. Variables:

1. Independent--

a. In analyzing six child behaviors, 28 instructional variables were used as predictor variables. These were taken directly from the Classroom Observation Instrument or constructed from items on that instrument. The 28 variables were (p. 238):

Variables	
No.	Name
16	Child/adult ratio (Number of children over teachers, aides, and volunteers)
24	Child selection of seating and work groups
28	Instructional materials used
38	Audio visual equipment used
40	General equipment, materials used
45	Total number different resource categories coded "used today" over three days
82	Wide variety of activities, concurrent
83	Wide variety of activities, over one day
104	Adult with one child
236	TV/Academic Activities
237	Audio visual equipment/Academic Activities
238	Exploratory materials/Academic Activities
239	Math or science equipment/Academic Activities
240	Texts, workbooks/Academic Activities
241	Puzzles, games/Academic Activities
357a	All adult questions to children
364a	Adult responses to child requests or questions, academic
365a	Adult responses to child requests or questions, nonacademic
375a	Adult instructs an individual child
389a	Adult general comments to children
390a	Adult task-related comments to children
394a	All adult acknowledgment to children
398a	All adult praise to children
438a	Adult communication or attention focus, one child
439a	Adult communication or attention focus, two children
440a	Adult communication or attention focus, small group
451a	Adult academic commands, requests, and direct questions to children
452a	Adult open-ended questions to children

Table 101

NUMBER OF CLASSROOMS AND SITES INCLUDED IN THE PARTIAL CORRELATIONS AND REGRESSION ANALYSES BY SPONSOR AND GRADE LEVEL*

Sponsor	First Grade		Third Grade	
	Number of Classrooms	Number of Sites Represented	Number of Classrooms	Number of Sites epresented
Far West Labs	12	3	14	4
University of Arizona	14	4	2	1
Bank Street College	11	4	7	2
University of Oregon	5	2	4	1
University of Kansas	17	5	12	4
High/Scope	13	4	0	0
EDC	12	3	6	2
Non-Follow Through	24	-	13	-
Total number of classrooms	108		58	

---

*All classrooms are in the entering kindergarten (EK) grade level.

p. 234

b. In analyzing absence and test outcomes, an additional 312 instructional variables--340 in all--were used as predictor variables. The report does not provide a comprehensive list of the 340. Instead, it reports those involved in significant associations with at least one of the criterion variables noted below.

2. Dependent--

a. Six child behaviors were used as criterion variables with reference to the smaller set of 28 instructional process variables mentioned above:

(1) Independence "defined as a child or a group of children engaged in any task without an adult" (p. 237).

(2) Task persistence "defined as a child engaged in self-instruction over a designated period of time (a matter of five interaction frames)" (p. 241).

(3) Cooperation "defined as two or more children working together on any joint task without an adult" (p. 242).

(4) Verbal initiative measured by noting "whenever a child asked questions, instructed someone else, engaged in a task-related conversation, made general comments, or provided feedback to other children" (p. 243).

(5) Observed Self-Esteem "operationally defined as existing where a child offers his opinion--he may praise or acknowledge the work of another child. This child also makes statements about his self-worth and extends his response when asked questions" (p. 244).

(6) Question Asking which is not defined. The commentary indirectly indicates that the term refers to the number of incidents in which a child asks a question, but further details, e.g., of whom, cannot be determined.

"Question asking, verbal initiative, and self-esteem were analyzed for first grade only since too few significant correlations were obtained in the third grade" (p. 237).

b. Six other variables were defined as criterion variables in reference to the larger set (340) of instructional process variables mentioned above:

(1) Days absent

(2) "Raven's Coloured Progressive Matrices (Raven's)

This test was originally designed by John C. Raven (1956, 1962) as a 'culture-fair,' nonverbal, intelligence test, but SRI uses the test as a measure of a child's problem-solving ability in visual perceptual tasks. Each test item presents the child with a pattern from which a piece is missing and the child is asked to select the appropriate piece from several alternatives. SRI prepared and uses an abbreviated 27-item, group-administered version of the original Raven's. Instructions were adapted by SRI from the 1965 Matrices Guide. These instructions were reviewed and accepted by agents designated by the author" (pp. 46-47).

(3) Intellectual Achievement Responsibility Scale (IAR)

"The IAR (Crandall, Katkovsky, & Crandall, 1965) is a group-administered, noncognitive measure designed to assess the extent to which the child takes responsibility for himself or attributes his successes or failures to the operation of internal or external forces. Each test item describes a positive or negative achievement experience and two alternative explanations of the event, one of which denotes internal control while the other denotes external control. The child is asked to select a response that describes whether he would be responsible for a particular event or someone else would be responsible for the event, e.g., I read well--because I study hard or because the books are easy.

"Because the original 34-item IAR was designed for older children, SRI rewrote the items in more simplistic terms to make it more appropriate for Follow Through third graders. Two subscores are derived: one for those items related to a child's feeling of success, and the other for the items related to a child's feelings of failure. The items are mutually exclusive and a child could score high on both scales. For both subscores, the higher the score, the greater the indication of a child's attributing his successes and failures to internal forces" (p. 47).

(4) Coopersmith Self-Esteem Inventory (Coopersmith)

"The Coopersmith is a noncognitive instrument designed by Stanley Coopersmith (1967) to assess a child's feelings of self-esteem. The test measures a child's feelings about himself, the way he thinks other

> people feel about him, and his feelings about school. The test contains 58 items, each of which consists of a set of statements. The child is asked to select which of the statements is 'like me' or 'not like me.' The SRI version of the Coopersmith test was group administered in Spring 1973 and contains the original 58 items of the inventory. However, since eight of these items are used merely as a validity check, the raw score for the Coopersmith is based on the 50 items designed to measure the child's self-esteem" (p. 47).

(5&6) Metropolitan Achievement Tests (MAT) Form F

> "These tests (by Durost et al., 1970) were first included in the national Follow Through evaluation in Spring 1972. They were selected because they cover several areas of achievement, have norms based on a standardization sample that includes children of low-income families, and are reliable. The subscores from the MAT Primary I Test for word analysis, total reading, and math were computed for the first grade children. Subscores from the MAT Elementary Test for total reading, language, math computation, math concepts, math problem solving, and total math were computed for the third grade children" (p. 46).

Reading and math scores on MAT were analyzed separately. The Raven's, IAR, and Coopersmith's tests were administered only to third graders.

Although Coopersmith's Self-Esteem Inventory was specified as a criterion variable, results of the partial correlation analysis of its relationship with instructional process variables are not reported.

3. Covariate: The influence of entry WRAT scores is partialled out of the relationship between predictor and criterion variables. The report describes WRAT in the following manner.

> "The WRAT--highly correlated with the MAT--is the baseline test for the study because it is the only test that was administered consistently during the first three years of Follow Through. The WRAT was initially used in the Follow Through evaluation because it is a short, multi-level, single instrument achievement test that measures achievement in reading, spelling, and arithmetic. The 1965 version of the WRAT test (J. F. Jastak & S. R. Jastak) is standardized and norms have been established, but the appropriateness of these norms for the Follow Through sample has been questioned. For this reason, the MAT, which has been standardized on a larger population, and which examines achievement in greater

depth at each grade level, was substituted for the WRAT as a measure of achievement in the Spring of 1972" (p. 46).

D. Analytical Method: First order partial correlations were undertaken on first grade and third grade data. Within those distinctions all classroom scores were pooled--i.e., FT and NFT classroom scores were lumped together, ". . . in order to investigate instructional processes and their relationship to child outcomes regardless of where they occurred" (p. 233).

The six child behaviors were correlated separately with each of the small set of 28 instructional process variables adjusting for entry WRAT scores. Days absent and the five test scores were correlated separately with each of the larger set of 340 instructional process variables adjusting for entry WRAT scores.

"Since many of the process variables [which] are correlated with the child outcomes do describe components of the models . . . we have also reported sponsor mean scores on the outcome variables" (p. 233). Total FT means and standard deviations are also reported alongside means and standard deviations of all NFT classrooms. No significance testing of these differences is undertaken.

E. Comments:

"The problems of interpreting correlational data need to be mentioned. They concern such pitfalls as capitalizing on chance results and drawing invalid inferences from the correlations. The problem of capitalizing on chance is compounded by the large number of tests of significance that were computed. Of the 340 partial correlations computed with data on child absence and test scores, approximately 17 significant correlations would be expected to occur by chance if the variables were statistically independent. In the correlations on child absence, there were 56 first grade and 65 third grade correlations that were found to be significant. The problem is that there is no way of estimating which correlations occurred by chance. Variables which correlate similarly at both first and third grades probably can be interpreted with more confidence. The strongest type of statement that can be made when interpreting these correlations takes the form: '[I]n classrooms where teachers were observed to be more positive toward children, the absence rate of children was lower.' In all the results to be reported, it must be understood that correlations are compared between class averages. Inferences about the meaning of correlations at the individual student level are unwarranted.

"We cannot infer causal relationships based on correlational findings. We adjusted for the baseline WRAT scores hoping to eliminate some of the spurious relationships; e.g., classes that have, on the average, a larger proportion of bright children may tend to have a higher average socioeconomic status, have fewer absences, and have higher test scores. Thus, the partial correlation was computed in an attempt to remove associations between process and outcome that might be due to the general level of children in a particular classroom. The results from these studies must be considered exploratory and useful primarily for generating hypotheses for other studies.

"Special caution is needed in interpreting the relations between the instructional variables and child behaviors discussed in this section. The two sets of data were obtained from classroom observations, but on separate days. However, some correlations may be due to the interdependence of the variables. For example, the child independence rate may be lower in situations where the instructional process requires frequent intervention by adults. Such program differences will most certainly affect the occurrence of the child behavior" (p. 236).

* * * * * * * * * *

"The results may be viewed as a source of hypotheses for future studies that focus on the child behaviors evaluated in this section" (p. 237).

In addition, it should be noted that child behaviors can be considered as process variables in their own right--not so much the product of the FT experience as the substance of that experience which leads to intended outcomes. The report does not indicate precisely how each child behavior was measured. The single reference to this aspect of the methodology refers to procedures that were identical to the observation of classrooms which generated data for the assessment of implementation analysis reviewed in the first summary. The reference is too scant to warrant a definitive statement, but it appears that the child behaviors referred to as outcomes in this analysis may have been taken directly from the Classroom Observation Instrument or constructed from variables on that instrument. For example, Variables 114 to 118 of the Classroom Check List plus many of the variables on the Five Minute Observation protocol could be used to construct an Independence score. If this line of speculation is correct, that portion of the present analysis which deals with independence, task persistence, cooperation, verbal initiative, and question-asking might be more appropriately titled INSTRUCTIONAL PROCESSES AS RELATED TO INSTRUCTIONAL PROCESSES.

The one exception to this observation is Observed Self-Esteem which, depending upon how it was actually measured, appears to be more of an outcome variable. However, as indicated, self-esteem was analyzed only for the first grade since "too few correlations were obtained in the third grade" (p. 237). That being the case, it is difficult to understand why Coopersmith's Self-Esteem Inventory isn't used--either instead of the SRI self-esteem measure or in its own right.

The report does not indicate the exact manner in which entry WRAT scores were taken into consideration. However, since the criterion for selection of a classroom was that at least 20% of the students have baseline data, there is no assurance that the percentage with baseline data is the same in all the classes. On the contrary, it is likely that it is not. The question then arises: Does the mean of 20% of the students in a class reflect the same underlying concept as the mean of 40%, 60%, or other percentages? If it does not, then it is unclear what is being controlled for when class averages of entry WRAT scores are held constant.

F. Results: The significant correlations identified in the analysis dealing with child behaviors are reported in Table 103 below. The particulars of this table cannot be adequately commented upon here. Readers are referred to the report for a more detailed discussion.

The significant correlations identified in the analysis dealing with children's absence and test scores fill over 27 pages in the report. We reproduce here the report's conclusions concerning each criterion variable followed by each sponsor's, total FT, and total NFT means and standard deviations for each variable when possible.

1. Absence--

"These data suggest that in both first and third grade classrooms, children may be absent less frequently in classrooms where there is a higher rate of child independence, child questioning, adults responding, individualized instruction, and open-ended questioning. Also in classrooms where children and adults show more positive affect (smiling or laughing) the children may be absent less often.

"Children were absent more often in both first and third grade classrooms where they worked in large groups more often and where adults used direct questions in academic work and frequent corrective feedback more often.

"Findings for the third grade indicate that in classrooms where children were punished more often, they also were absent more often. In addition, classrooms with a higher rate of negative affect on the part of teachers and students showed a higher absence rate.

"A report of sponsor absence rate by grade level indicates that at the first grade level, those sponsors who used more highly structured environments, materials, and interactions also had a higher absence rate (see Table 106). Two sponsors, Far West Labs and University of Arizona, who used a wide variety of activities and materials and have classrooms

Table 103

PARTIAL CORRELATIONS OF INSTRUCTIONAL VARIABLES AND CHILD BEHAVIORS
FIRST (N=105) AND THIRD (N=55) GRADES
(Follow Through and Non-Follow Through)

Variables		Independence				Task Persistence				Cooperation			
		1st Grade		3rd Grade		1st Grade		3rd Grade		1st Grade		3rd Grade	
No.	Name	r	p<	r	p<	r	p<	r	p<	r	p<	r	p<
16	Child/adult ratio (Number of children over teachers, aides, and volunteers)	.23	.05	.06		.09		-.15		.02		-.10	
26	Child selection of seating and work groups	.36	.001	.33	.01	-.22	.05	-.08		.19	.05	.35	.01
28	Instructional materials used	-.01		-.03		.11		-.02		.09		.23	.10
38	Audio visual equipment used	.13		.12		-.25	.01	-.28	.05	.15		.28	.05
40	General equipment, materials used	.22	.05	.03		-.09		-.02		.09		.08	
65	Total number different resource categories coded "used today" over three days	.13		.12		-.23	.05	-.25	.10	.18	.10	.37	.01
82	Wide variety of activities, concurrent	.22	.05	.08		-.12		-.03		.15		.13	
83	Wide variety of activities, over one day	.43	.001	.37	.01	-.36	.001	-.41	.01	.32	.01	.39	.01
104	Adult with one child	.57	.001	.46	.00	-.16		-.27	.05	.08		.31	.05
236	TV/Academic Activities	-.03		-.11		-.10		-.09		-.11		.10	
237	Audio visual equipment/Academic Activities	.24	.01	.28	.05	-.25	.01	-.18		-.01		-.07	
238	Exploratory materials/Academic Activities	.34	.001	.01		-.22	.05	-.26	.10	.27	.01	.37	.01
239	Math or science equipment/Academic Activities	-.18	.10	-.29	.05	.17	.10	-.04		-.18	.10	-.18	
240	Texts, workbooks/Academic Activities	-.33	.001	-.34	.01	.31	.01	.57	.001	-.49	.001	-.48	.001
241	Puzzles, games/Academic Activities	.16		.12		-.07		-.04		.09		.25	.10
357a	All adult questions to children	-.17	.10	-.17		.03		-.27	.05	-.17	.10	-.14	
364a	Adult responses to child requests or questions, academic	.13		.31	.05	-.15		.34	.01	.06		-.02	
365a	Adult responses to child requests or questions, nonacademic	.42	.001	.44	.001	-.33	.001	-.31	.05	.21	.05	.49	.001
375a	Adult instructs an individual child	-.09		.32	.05	.23	.05	-.06		-.17	.10	.17	
389a	Adult general comments to children	.22	.05	.09		-.12		-.24	.10	-.13		-.04	
390a	Adult task-related comments to children	.12		.25	.10	-.24	.05	.44	.001	-.39	.001	.31	.05
394a	All adult acknowledgment to children	-.16		-.47	.001	.15		.31	.05	-.11		-.30	.05
398a	All adult praise to children	-.60	.001	-.22		.20	.05	-.41	.01	-.21	.05	-.13	
438a	Adult communication or attention focus, one child	-.01		.02		.13		-.30	.05	-.06		.02	
439a	Adult communication or attention focus, two children	.29	.01	.02		-.13		.02		.28	.01	.19	
440a	Adult communication or attention focus, small group	-.15		-.20		.19	.05	.05		.01		-.09	
451a	Adult academic commands, requests, and direct questions to children	-.41	.001	-.30	.05	.01		-.35	.01	-.28	.01	-.28	.05
452a	Adult open-ended questions to children	.16		.29	.05	-.12		-.10		.13		.29	.05

Note: Underscored figures are significant.

.05 = 5 chances in 100 that the relationship would occur by chance.
.01 = 1 chance in 100 that the relationship would occur by chance.
.001 = 1 chance in 1,000 that the relationship would occur by chance.

Initiative				Self-Esteem		Child Questions	
1st Grade		3rd Grade		1st Grade		1st Grade	
r	p[c]	r	p[c]	r	p[c]	r	p[c]
-.17	.10	-.22	.10	-.11		-.15	
.11		.27	.05	.14		.03	
-.24	.01	.13		-.08		-.07	
-.01		.30	.05	-.03		-.12	
.06		.23	.10	.04		.005	
.07		.36	.01	-.05		.03	
.22	.05	-.10		.22	.05	.09	
.22	.05	.15		-.08		.14	
.08		.55	.001	.03		.14	
-.03		.19		-.02		-.03	
.06		.28	.05	.06		-.04	
.05		.23	.10	.28	.01	-.11	
-.08		-.02		.05		.11	
-.09		-.31	.05	.07		-.04	
.04		-.002		-.09		-.07	
.50		-.16		.24	.05	-.04	
.10		-.15		.10		.23	.05
.31	.01	.39	.01	-.10		.27	.01
.09		.43	.001	-.05		.22	.05
.25	.01	.39	.01	-.10		.36	.001
-.09		.25	.10	-.09		-.16	.10
-.13		-.24	.10	-.30	.01	.04	
-.17	.10	.02		.04		.02	
-.01		.20		-.24	.01	.38	.001
.05		-.05		.42	.001	-.03	
.04		-.13		.48	.001	-.32	.001
-.04		-.37	.01	.21	.05	.03	
.10		.18		-.07		-.07	

Table 106

MEANS AND STANDARD DEVIATIONS OF DAYS ABSENT FOR SPONSORS AND NON-FOLLOW THROUGH

	First Grade*		Third Grade*	
	Mean	Standard Deviation	Mean	Standard Deviation
Far West Lab	9.81	1.80	7.99	1.40
University of Arizona	12.37	4.34	6.34	0.78
Bank Street College	14.90	6.41	9.64	6.42
University of Oregon	17.12	3.72	12.12	1.63
University of Kansas	15.34	16.70	10.28	2.77
High/Scope	13.36	3.25	—	—
EDC	14.25	10.90	10.53	7.46
All Follow Through	14.15	6.33	9.50	4.30
Non-Follow Through	14.05	5.65	9.59	2.59

---

*Adjusted scores.

where children exhibited independent behavior, had the lowest absence rate of all sponsors at both the first and third grade levels. The data on Table 106 also indicate that the absence rate for all sponsors and Non-Follow Through diminished from first grade to third grade. While the means and standard deviations of Follow Through and Non-Follow Through were very similar, Far West's and University of Arizona's absence rates were considerably lower than Non-Follow Through rates" (pp. 255-256).

2. Raven's Coloured Progressive Matrices--

   The results of the ". . . partial correlation analysis . . . support the notion that children are likely to do better in perceptual problem solving tests if they are in classrooms that use instructional processes similar to those in the more open models [i.e., Far West, University of Arizona, Bank Street, and EDC]" (p. 265).

3. IAR Scale--

   "Findings from this study seem to indicate that classrooms that allow more independence have children who accept responsibility for their success. However, classrooms that allow more child involvement in selecting their groups, that allow child independence in activities, that provide individualized instruction, and that have a wide variety of activities and materials, do not as often have children who accept responsibility for their own failure.

   "On the other hand, children who are in classrooms that are more teacher-directed--and who experience education more often in large groups--are more likely to attribute their success to the teacher or to easy work. The failure of children in the more structured, teacher-directed classrooms is accepted by the child as his own failure.

   "Interestingly, the Non-Follow Through classrooms had positive gain scores on both the success and failure scores (see Table 10).

   "In fact, children in Non-Follow Through classrooms took more responsibility for their failure than children in any of the sponsors' classrooms" (pp. 274-275).

4. MAT-Math--

   Math scores were positively related to the opportunity children had to engage in math activity, measured in terms of both time spent on math as well as the length of the school day.

Table 108

MEANS AND STANDARD DEVIATIONS ON THE RAVEN'S COLOURED PROGRESSIVE MATRICES TEST FOR SPONSORS* AND NON-FOLLOW THROUGH

	Var. 440/Fall 1969 WRAT		Var. 448/Raven's Raw Score		Relative Difference in Adjusted Raven's Scores Among Sponsors and NFT*	
Sponsor and Number	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
Far West (N=14)	40.78	3.31	21.16	1.81	.42	1.69
University of Arizona (N=2)	28.20	.28	19.83	1.65	1.69	1.59
Bank Street (N=7)	32.42	6.78	19.54	2.56	.63	1.65
University of Kansas (N=4)	33.50	6.37	18.39	2.92	-.84	1.95
University of Oregon (N=12)	35.16	5.82	17.59	2.37	-1.99	1.42
EDC (N=6)	26.38	6.95	19.02	1.09	1.25	1.71
Non-Follow Through (N=13)	38.75	8.71	20.74	2.12	.52	1.40

---

*Too few High/Scope third grade classrooms were both tested and observed to permit including the High/Scope model in the analysis (p. 264).

Table 110

SPONSOR AND NON-FOLLOW THROUGH MEANS AND STANDARD DEVIATIONS ON THE INTELLECTUAL ACHIEVEMENT SCALE

	Var. 440/Fall 1969 WRAT		IAR - Success (raw score)		IAR - Failure (raw score)	
Sponsor and Number	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
Far West (N=14)	40.78	3.31	13.54	1.08	9.23	.93
University of Arizona (N=2)	28.20	.28	12.65	1.20	8.48	1.86
Bank Street (N=7)	32.42	6.78	13.04	.64	9.20	1.21
University of Kansas (N=4)	33.50	6.37	13.12	.82	10.07	1.31
University of Oregon (N=12)	35.16	5.82	12.82	.58	10.14	.54
High/Scope (N=0)	*	*	*	*	*	*
EDC (N=6)	26.38	6.95	13.14	1.17	9.61	1.17
NFT (N=13)	38.25	8.71	13.53	1.02	10.82	1.12

---

*These adjusted scores indicate that relative to all classrooms in the sample these sponsors entering ability is parceled out. This is identified as a residual gain score. The mean of is zero; therefore, some sponsors will have a minus score relative to all others.

p. 269

Relative Difference in Adjusted IAR Scores Among Sponsors and Non-Follow Through

Success (gain score)		Failure (gain score)	
Mean	Standard Deviation	Mean	Standard Deviation
.13	1.03	-.80	.92
-.07	1.22	-1.08	1.85
.01	.52	.52	1.14
-.10	.60	.30	1.12
-.38	.46	1.12	.31
*	*	*	*
.41	1.03	.11	.99
.16	.97	.88	1.13

have gained when
the total classrooms

When children are divided into three groups--high, medium, and low--on the basis of their entry WRAT scores, an interesting finding was obtained. The correlation between time spent on math and math scores was high for students in the low group and low for students in the high group. The report ". . . suggests that children who entered school with less ability need more time spent in developing basic skills.

"The occurrence of reading and the number of children involved in reading were related to math scores. This relationship suggests that programs that stress one of the basic skills also stress the other basic skills" (p. 283).

5. MAT-Reading--

Overall conclusions are not provided. A few highlights of the findings are that MAT-reading scores are positively correlated with:

-- the amount of time spent in school,
-- the amount of time spent in reading,
-- the amount of time spent in other studies which required reading skills.

There was some indication that younger children do better on reading when adults work with smaller groups (3 to 8 children), but older children do better when adults work with larger groups (9 or more). Reading scores were negatively correlated with both "a wide variety of materials" and "a wide variety of activities" in the classroom.

Table 114

SPONSOR AND NON-FOLLOW THROUGH MEANS AND STANDARD DEVIATIONS FOR MAT TOTAL MATH SCORES

Sponsor and Number	Var. 438/Fall 1971 WRAT		Var. 443/ MAT Math Raw Score		Relative Difference in adjusted MAT Math Scores Among Sponsors and NFT	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
			FIRST GRADE			
Far West (N=12)	32.65	2.78	31.31	5.73	-2.56	6.01
University of Arizona (N=14)	30.92	5.09	31.08	7.26	-1.55	6.14
Bank Street (N=11)	26.86	5.72	29.54	5.22	-.22	4.50
University of Oregon (N=5)	29.32	7.50	31.14	7.25	-.37	3.10
University of Kansas (N=17)	28.11	4.68	36.76	7.50	6.10	7.36
High/Scope (N=13)	26.60	7.10	24.24	7.03	-5.34	5.66
EDC (N=12)	27.02	5.40	30.99	4.00	1.11	1.67
NFT (N=24)	32.67	9.05	34.27	9.99	.38	6.28
			THIRD GRADE			
Far West (N=14)	40.78	3.31	49.62	9.71	-6.08	8.65
University of Arizona (N=2)	28.20	.28	33.72	7.24	-12.54	7.45
Bank Street (N=7)	32.42	6.78	47.42	12.49	-2.00	8.42
University of Oregon (N=4)	33.50	6.37	71.12	9.32	20.89	8.13
University of Kansas (N=12)	35.16	5.82	52.63	10.67	1.15	8.18
High/Scope (N=0)	*	*	*	*	*	*
EDC (N=6)	26.38	6.95	38.53	11.56	-6.36	7.20
NFT (N=13)	38.25	8.71	58.79	11.96	5.00	14.43

---

*Insufficient classrooms were both tested and observed in the third grade for High/Scope for their model to be included in the analysis (p. 290).

Table 116

SPONSOR AND NON-FOLLOW THROUGH MEANS AND STANDARD DEVIATIONS FOR MAT READING SCORES

Sponsor and Number	Var. 438/Fall 1971 WRAT		Var. 442/ MAT Reading Raw Score		Relative Difference in adjusted MAT Reading Scores Among Sponsors and NFT	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
FIRST GRADE						
Far West (N=12)	32.65	2.78	40.81	7.16	-5.30	7.30
University of Arizona (N=14)	30.92	5.09	41.99	8.95	-1.55	5.48
Bank Street (N=11)	26.86	5.72	36.29	7.18	-2.53	2.49
University of Oregon (N=5)	29.32	7.50	48.61	9.94	6.93	3.19
University of Kansas (N=17)	28.11	4.68	46.13	9.15	5.86	9.15
High/Scope (N=13)	26.60	7.10	36.00	8.55	-2.52	6.24
EDC (N=12)	27.02	5.40	34.42	6.60	-4.58	4.48
NFT (N=24)	32.67	9.05	48.34	14.02	2.77	7.73
THIRD GRADE						
Far West (N=14)	40.78	3.31	47.20	8.40	-1.78	6.87
University of Arizona (N=2)	28.20	.28	32.95	2.05	-3.83	1.78
Bank Street (N=7)	32.42	6.78	35.34	11.44	-5.53	5.50
University of Oregon (N=4)	33.50	6.37	46.63	11.70	4.71	6.04
University of Kansas (N=12)	35.16	5.82	42.21	9.99	-1.32	8.03
High/Scope (N=0)	*	*	*	*	*	*
EDC (N=6)	26.38	6.95	32.09	10.12	-2.93	7.03
NFT (N=13)	38.25	8.71	53.15	5.39	6.62	9.40

---

*Insufficient classrooms were both tested and observed in the third grade for High/Scope for their model to be included in the analysis.

IV. INSTRUCTIONAL PROCESSES AS RELATED TO CHILD OUTCOMES--MULTIPLE REGRESSION (pp. 307-316)

This summary is the second of three summarized analyses that looked at associations between selected process and outcome variables.

A. Research Questions:

1. How much variance in child outcome scores is explained by entry WRAT scores?
2. How much variance, over and above that explained by entry WRAT scores, can be explained by a set of selected instructional process variables?
3. How much variance, over and above that explained by entry WRAT scores, is explained by respective instructional process variables taken individually?

B. Sample: The sample appears to be the same as that used in the partial correlation analyses reported earlier--i.e., 108 first grade and 58 third grade classes.*

C. Variables: Generally, the same variables used in the partial correlation analysis summarized earlier are used here. However, the report notes that "a set of process variables was chosen separately for each outcome score based on the partial correlations in the first part of the section" (p. 310). Presumably, those with the largest first-order, partial correlations were selected.

D. Analytical Method:

"A stepwise multiple linear regression was run separately for each grade level with each outcome score as the dependent variable and the set of selected process variables and the baseline WRAT score as the independent variables. The baseline WRAT score was entered into the regression on the first step. The process variables were added one at a time until a total of 20 variables had been included. Of particular interest in examining the relationship of the set of process variables to the outcome variables is the degree to which the set of process variables 'predicted' the outcome after the effect of the baseline WRAT had been accounted for. The coefficient of determination, R^2, was used to this end. Since the baseline WRAT entered the regression at the first step, the R^2 on this step was a measure of the degree to which the baseline WRAT score accounted for the variance in outcome score. The difference between the value of R^2 at any subsequent stage and the R^2 at the first step was a measure of the degree to which the process variables provided a prediction of the outcome scores beyond that available from the baseline WRAT alone. The significance of the contribution of any variable that is added to the regression is indicated by the 'f-to-enter' statistic. This statistic corresponds to the test of whether the variable, when it entered the regression, had a partial regression coefficient significantly different from zero. Variables were added until no further significant increases in R^2 occurred" (p. 310).

---

*The n's reported in the results tables indicate that three first grade classes and one third grade class were lost somewhere along the line.

E. Comments: Acknowledging the shortcomings of the analysis, its results are the strongest evidence to be found among the FT evaluations reviewed that particular classroom activities are related to specified outcomes. We recommend the details of Table 123 presented below for careful consideration.

Given the potential for interpretation that the results of this analysis has, it is unfortunate that the report does not specify exactly which instructional process variables were selected for the regression of each outcome variable. Similarly, it's unfortunate that the decision rules for the selection of instructional process variables are not made clear.

The manner in which process variables were selected is clouded by lack of specification. The manner in which selected variables are introduced in the multiple regression routine is also unclear. At one point the report states that "the process variables were added one at a time until a total of 20 variables had been included" (p. 310). Later, the report states that "variables were added until no further significant increases in R^2 occurred" (p. 310).

As engaging as the results are, they do not provide conclusive evidence on which to base causal inferences. The overall research design precludes definitive conclusions concerning causal relationships between the examined process and outcome variables.

F. Results: The number of process variables selected for each outcome measure, WRAT's contribution to R^2, total R^2 explained by retained process variables, the number of retained process variables are reported in Table 122 reproduced below. Of particular interest are:

1. In and of themselves, WRAT scores explain a small percentage of the variation in child behavior scores. This contrasts sharply with the amount of variance explained by most of the process variables retained when child behaviors are regressed on them.

2. The relatively few number of process variables that are retained ("included") in the regression, which implies the overall high level of inter-collinearity among selected variables; and

3. The fairly "respectable" R^2 values, especially for academic test scores and especially for third grade outcome measures.

The particular process variables retained in the regressions of child behavior and test score (cum absence), outcome measures are reported in Table 123. Also reported are increments to R^2 attributable to each variable.

The report makes the following general observations:

"In the test outcome domain, the contributions of the process variables were substantial across the board. For the MAT reading and math outcome scores, the results for first and third grade are very similar. The process variables added more than .40 to the R^2 value for MAT math and more than .25 to the R^2 value for MAT reading. For both grade levels the R^2 at the last step corresponded to a multiple correlation between .85 and .90. For the Raven's, the COI variables incremented the R^2 value by more than the baseline WRAT. This indicates that the classroom processes may be as important a factor as initial ability in accounting for performance on the Raven's. The results on the IAR Failure scale are also amazing. The baseline WRAT accounted for only 4% of the variability, and, when the COI variables were added, the coefficient of determination was .83. This corresponds to a multiple correlation coefficient of over .90. Though such values are clearly overestimates, the result provides compelling evidence for a strong relationship between the way children take responsibility for their failures and the classroom processes to which they are exposed.

"The number of process variables included in the regressions before the 'F-to-enter' dropped below the .05 level was a minimum of two, for Child Questions, and a maximum of 14, for days absent at the first grade level (see Table 122). For the most part, the number of process variables that entered the regression was between 2 and 6 for the behavioral outcomes and between 7 and 11 for the outcome test scores. The number of the variables that entered any particular regression relative to the number of process variables included is an indication of the high degree of dependency among the process variables" (p. 312).

Table 122

SUMMARY STATISTICS FOR THE STEPWISE REGRESSION ANALYSES

	FIRST GRADE (N=105)					
Outcome Variable	Number of Selected Process Variables	R^2_{WRAT}	$R^2_{WRAT,}$ Process Variables	Multiple R	Number of Process Variables (included in regression)[a]	Number of Selected Process Variables
Behavioral Outcome Variables						
Child Questions	27	.00	.28	.53	2	27
Self-Esteem	27	.00	.48	.69	5	27
Child Independence	27	.00	.67	.82	5	27
Task Persistence	27	.00	.44	.66	9	27
Cooperation	27	.00	.32	.57	2	27
Verbal Initiative	27	.00	.22	.47	3	27
Days Absent	38	.01	.67	.82	14	38
Test Outcome Variables						
MAT Math	75	.32	.74	.86	10	75
MAT Reading	67	.50	.73	.85	8	66
Ravens	—	—	—	—	—	97
IAR-Success	—	—	—	—	—	18
IAR-Failure	—	—	—	—	—	85

---

[a] This column contains the number of process variables that entered the stepwise regression with at .05.

THIRD GRADE (N=58)

R^2_{WRAT}	R^2_{WRAT}, Process Variables	Multiple R	Number of Process Variables (included in regression)*
.00	.29	.54	3
.01	.42	.65	6
.01	.41	.64	3
.06	.61	.78	6
.10	.61	.78	5
.00	.38	.62	2
.07	.69	.83	6
.17	.81	.90	8
.42	.79	.89	7
.41	.86	.93	9
.18	.57	.75	4
.04	.83	.91	11

an "F-to-enter" that was significant

Table 123

STATISTICS FROM THE STEPWISE REGRESSION OF THE BASELINE WRAT AND SELECTED COI VARIABLES ON THE OUTCOME SCORES

Step	Variable No.	Variable Name	Multiple R	R^2	RSQ Change	F-to-Enter
Dependent Variable: Child Questions						
First Grade (N=105)						
1		F71 WRAT	.025	.001	.001	.067
2	438a	Adult communication or attention focus, child	.385	.148	.147	17.641
3	389a	Adult general comments to children	.528	.279	.131	18.359
Third Grade (N=57)						
1		F69 WRAT	.056	.003	.003	.176
2	389a	Adult general comments to children	.366	.134	.131	8.169
3	236	TV	.499	.250	.115	8.145
4	398a	All adult praise to children	.540	.292	.042	3.103
Dependent Variable: Self-Esteem						
First Grade (N=105)						
1		F71 WRAT	.045	.002	.002	.213
2	440a	Adult communication or attention focus, small group	.477	.228	.226	29.815
3	439a	Adult communication or attention focus, two children	.581	.337	.109	16.680
4	394a	All adult acknowledgment to children	.627	.393	.056	9.147
5	451a	Adult academic commands, requests, and direct questions to children	.666	.443	.050	8.946
6	104	Adult with one child	.696	.484	.041	7.816
Third Grade (N=57)						
1		F69 WRAT	.090	.008	.008	.449
2	390a	Adult task-related comments to children	.291	.084	.076	4.503
3	389a	Adult general comments to children	.401	.161	.077	4.848
4	240	Texts, workbooks	.474	.224	.063	4.232
5	83	Wide variety of activities over one day	.536	.287	.063	4.486
6	24	Child selection	.604	.365	.078	6.110
7	394a	All adult acknowledgment to children	.646	.418	.053	4.462
Dependent Variable: Child Independence						
First Grade (N=105)						
1		F71 WRAT	.060	.004	.004	.373
2	398a	All adult praise to children	.598	.357	.354	56.158
3	104	Adult with one child	.732	.536	.178	38.782
4	16	Child/adult ratio	.779	.606	.070	17.862
5	82	Wide variety of activities; concurrent	.805	.649	.043	12.001
6	439a	Adult communication or attention focus, two children	.817	.668	.019	5.626
Third Grade (N=57)						
1		F69 WRAT	.094	.009	.009	.487
2	394a	All adult aknowledgment to children	.482	.232	.223	15.687
3	104	Adult with one child	.580	.336	.104	8.302
4	452a	Adult open-ended questions to children	.641	.411	.075	6.596
Dependent Variable: Task Persistence						
First Grade (N=105)						
1		F71 WRAT	.054	.003	.003	.306
2	83	Wide variety of activities over day	.367	.135	.132	15.529
3	375a	Adult instructs an individual child	.445	.198	.063	7.917
4	237	Audio visual equipment	.487	.237	.040	5.190
5	240	Tests, workbooks	.524	.275	.038	5.119

Table 123 (Continued)

STATISTICS FROM THE STEPWISE REGRESSION OF THE BASELINE WRAT AND SELECTED COI VARIABLES ON THE OUTCOME SCORES

Step	Variable No.	Variable Name	Multiple R	R^2	RSQ Change	F-to-Enter
Dependent Variable: Task Persistence						
First Grade (N=150) (Continued)						
6	451a	Adult academic command, request, and direct questions to children	.556	.309	.034	4.877
7	440a	Adult communication or attention focus, small group	.587	.344	.035	5.161
8	16	Child/adult ratio	.609	.371	.027	4.053
9	365a	Adult responses to child request or question, nonacademic	.633	.401	.030	4.794
10	438a	Adult communication or attention focus, one child	.661	.437	.036	6.054
Third Grade (N=57)						
1		F69 WRAT	.241	.058	.058	3.377
2	240	Texts, workbooks	.592	.350	.292	24.280
3	364a	Adult responses to child requests or questions, academic	.658	.432	.082	7.684
4	83	Wide variety of activities over one day	.698	.488	.055	5.621
5	16	Child/adult ratio	.725	.526	.038	4.085
6	104	Adult with one child	.756	.572	.046	5.421
7	239	Math or science equipment	.781	.610	0.38	4.746
Dependent Variable: Cooperation						
First Grade (N=105)						
1		F71 WRAT	.004	.000	.001	.128
2	240	Texts, workbooks	.491	.241	.241	32.759
3	390a	Adult task-related comments to children	.569	.324	.082	12.421
Third Grade (N=57)						
1		F69 WRAT	.317	.100	.100	6.144
2	240	Texts, workbooks	.571	.326	.226	18.082
3	45	Total different resource categories coded "used today" over three days	.646	.417	.091	8.271
4	40	General equipment, materials used	.723	.523	.106	11.535
5	82	Wide variety of activities, concurrent	.754	.569	.046	5.479
6	237	Audio visual equipment	.779	.606	.037	4.717
Dependent Variable: Verbal Initiative						
First Grade (N=105)						
1		F71 WRAT	.020	.000	.000	.040
2	365a	Adult responses to child request or question, nonacademic	.312	.098	.098	10.980
3	28	Instructional materials used	.379	.144	.046	5.467
4	45	Total different resource categories coded "used today" over three days	.463	.215	.071	9.032
Third Grade (N=57)						
1		F69 WRAT	.002	.000	.000	.151
2	104	Adult with one child	.497	.247	.247	18.033
3	451a	Adult academic commands, requests, and direct questions to children	.618	.382	.081	6.934

Table 123 (Continued)

STATISTICS FROM THE STEPWISE REGRESSION OF THE BASELINE WRAT AND SELECTED COI VARIABLES ON THE OUTCOME SCORES

Step	Variable No.	Variable Name	Multiple R	R^2	RSQ Change	F-to-Enter
Dependent Variable: Days Absent--Positive and Negative						
First Grade (N=105)						
1		F71 WRAT	.118	.014	.014	1.456
2	104	Adult with one child	.379	.144	.130	15.470
3	231	Volunteer involved	.475	.226	.082	10.700
4	596c	Adult feedback to children for behavior	.539	.290	.064	9.065
5	471a	Adults attentive to large group	.604	.365	.075	11.733
6	462a	All positive behavior	.638	.406	.041	6.786
7	85	Teacher without children	.667	.445	.039	6.765
8	256	Practical skills acquisition	.691	.478	.033	6.025
9	127	Large group with aide	.710	.504	.026	4.965
10	417a	Children attentive to adults, academic	.727	.529	.025	5.024
11	453a	Adult responds to child's question with a question	.754	.568	.039	8.349
12	353a	Adult commands, requests, and directs questions to groups of children, academic	.770	.593	.026	5.794
13	601c	All adult neutral corrective feedback	.788	.621	.027	6.543
14	585c	Child's extended response to question	.808	.653	.033	8.533
15	125	Two children with aide	.819	.670	.017	4.454
Third Grade (N=57)						
1		F69 WRAT	.272	.074	.074	4.485
2	463a	All negative behavior	.536	.288	.214	16.500
3	125	Two children with aide	.624	.390	.102	9.012
4	114	One child without adults	.710	.504	.115	12.260
5	371a	Child extensive response to adult open-ended question	.758	.575	.071	8.689
6	417a	Children attentive to adults, academic	.801	.642	.067	9.477
7	351a	Adult commands, requests, and directs questions to groups of children, nonacademic	.829	.687	.045	7.274
Dependent Variable: MAT Math						
First Grade (N=105)						
1		F71 WRAT	.594	.323	.353	56.095
2	412a	Adult feedback to children responding to adult academic command or question	.697	.485	.133	26.319
3	369a	Child extended response, nonacademic	.742	.550	.065	14.570
4	408a	Adult positive corrective feedback, academic	.771	.594	.044	10.790
5	441a	Adult communication or attention focus, group	.791	.625	.031	8.309
6	469a	All adult reinforcement with tokens	.809	.654	.029	8.197
7	370a	Child extended responses, academic	.825	.681	.027	8.074
8	354a	Adult communication, request, and direct question to individual child, academic	.837	.701	.020	6.386
9	76	Social interaction	.845	.714	.014	4.538
10	360a	Child responses, academic	.852	.727	.012	4.177
11	449a	Adult neutral corrective feedback, academic	.859	.739	.012	4.241
Third Grade (N=57)						
1		Classroom mean F69 WRAT	.410	.168	.168	11.096
2	582c	Adult academic command, request, and direct question to children	.642	.412	.245	22.480
3	135	Large group with any adults	.742	.551	.139	16.387
4	380a	Child self instructs, academic, with objects	.785	.616	.064	8.694
5	372a	Child presents information to a group	.818	.669	.053	8.199
6	66	Numbers, math, arithmetic	.848	.719	.050	8.871
7	24	Child selection	.872	.760	.042	8.484
8	76	Social interaction	.888	.788	.028	6.322
9	585c	Child's extended response to question	.898	.807	.019	4.620

Table 123 (Concluded)

STATISTICS FROM THE STEPWISE REGRESSION OF THE BASELINE WRAT AND SELECTED COI VARIABLES ON THE OUTCOME SCORES

Step	Variable No.	Variable Name	Multiple R	R^2	RSQ Change	F-to-Enter
Dependent Variable: MAT Reading						
First Grade: (N=105)						
1		F70 WRAT	.706	.501	.501	103.404
2	566c	Total academic verbal interaction	.791	.625	.124	33.888
3	163	Total weight in reading groups	.805	.648	.023	6.472
4	493c	Individual child response to adult academic command, request, or direct questions	.812	.660	.012	3.565
5	600c	All adult reinforcement with tokens	.825	.680	.020	6.131
6	370a	Child extended response, academic	.834	.696	.016	5.012
7	449a	Adult neutral corrective feedback, academic	.842	.709	.014	4.619
8	89	Teacher with large group	.849	.720	.011	3.721
9	107	Adult with large group	.856	.733	.012	4.415
Third Grade (N=57)						
1		F70 WRAT	.649	.421	.421	40.016
2	135	Large group with any adults	.763	.582	.161	20.753
3	15	Child/teacher and aide ratio	.787	.620	.038	5.245
4	566c	Total academic verbal interaction	.814	.663	.043	6.681
5	17	Total class duration	.837	.701	.038	6.479
6	157	Small group with any adults	.866	.751	.050	10.069
7	24	Child selection	.878	.772	.021	4.452
8	158	Large group with any adults	.889	.790	.018	4.076
Dependent Variable: Ravens Third Grade (N=57)						
1		F69 WRAT	.639	.408	.408	38.584
2	412a	Adult feedback to child response to adult academic command or question	.749	.561	.153	19.098
3	150	Large group with aide	.797	.635	.075	11.100
4	593c	All positive behavior	.835	.697	.067	10.809
5	524c	Child corrective feedback	.858	.736	.039	7.584
6	357a	All adult questions to children	.873	.761	.026	5.467
7	111	Small group of children with any adult	.891	.793	.032	7.699
8	185	Large group without adults	.905	.819	.026	6.903
9	44	Total number different resource categories coded "present" over three days	.919	.845	.026	8.120
10	519c	Child task-related comments to adults	.927	.858	.014	4.485
Dependent Variable: IAR - Success Third Grade (N=57)						
1		F69 WRAT	.429	.184	.184	12.416
2	79	Transitional activities	.594	.353	.169	14.117
3	184	Small group without adults	.657	.432	.078	7.319
4	145	Small group with teacher	.713	.508	.077	8.096
5	157	Small group with any adults	.755	.570	.062	7.317
Dependent Variable: IAR - Failure Third Grade (N=57)						
1		F69 WRAT	.200	.040	.040	2.285
2	138	Small group of children without adults	.599	.358	.319	26.818
3	517c	Child task-related comments to children	.698	.488	.129	13.385
4	15	Child/teacher and aide ratio	.745	.555	.067	7.792
5	515c	Small group working together using concrete objects	.775	.600	.045	5.774
6	366a	Adult responds to child question with open-ended question	.805	.648	.048	6.855
7	93	Aide with two children	.828	.685	.037	5.768
8	242	Percentage of CCLs on which an academic activity is occurring	.859	.738	.053	9.781
9	222	Small group of children with any adults	.878	.771	.032	6.622
10	350a	Child questions to adults	.890	.793	.022	4.895
11	135	Large group with any adults	.901	.812	.020	4.70[illegible]
12	159	One child without adults	.911	.830	.017	4.463

V. THE COOLEY-LOHNES MODEL: COMMONALITY ANALYSIS (pp. 317-325)

A. Research Questions: How much of the explained variance in academic performance can be attributed respectively to

-- entering ability
-- instructional processes
-- common variance between the two?

B. Sample: One hundred and twelve classrooms of Cohort III, entering kindergarten students, tested at the end of the first grade in Spring 1973 were selected from 25 diverse geographical locations. Details as to distribution of classrooms across sponsors are not provided in the report.

C. Variables:

1. Independent--Twenty measures were selected a priori from the Classroom Observation Inventory on the basis of their reflecting one of four components in a theoretical model of instructional processes--the Cooley-Lohnes model. The model is diagrammed in Figure 41; the measures selected to reflect each instructional process component are listed in Table 124. "The value of each component was found by summing the values of its process variables" (p. 319). In addition, Fall WRAT scores are used as measures of "initial pupil ability."

2. Dependent--"The sum of the MAT total Reading and Word Analysis scores was used to measure achievement ability" (p. 322).

D. Analytical Method:

"A commonality analysis approach was used to assess both the shared and the unique predictive power of each independent variable and of each set of independent variables. In commonality analysis, the variation of the dependent variable is divided into:

(1) The unique variance that may be predicted uniquely by each independent variable respectively;

(2) The shared variance that may be predicted by overlapping aspects of the independent variables; and

(3) The error variance, which is not predictable by the regression.

For the Cooley model, the independent variables were partitioned into two groups:

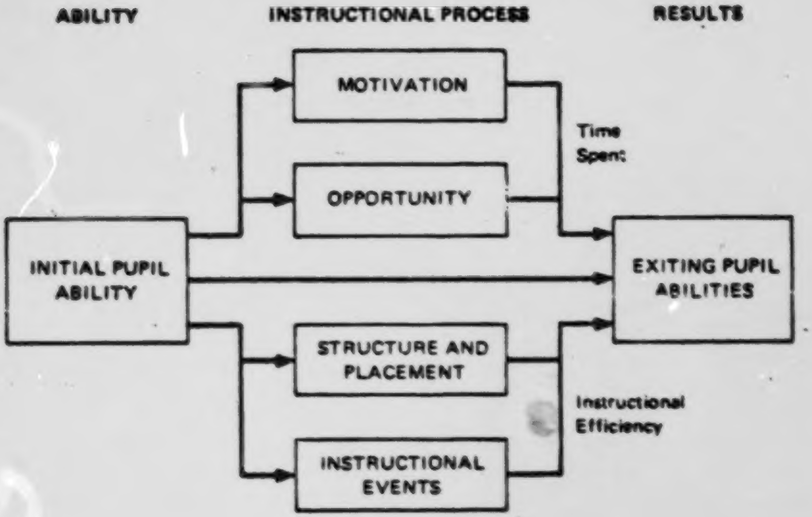

ABILITY
INSTRUCTIONAL PROCESS
RESULTS
MOTIVATION
OPPORTUNITY
Time Spent
INITIAL PUPIL ABILITY
EXITING PUPIL ABILITIES
STRUCTURE AND PLACEMENT
INSTRUCTIONAL EVENTS
Instructional Efficiency

FIGURE 41 DIAGRAM OF THE COOLEY AND LOHNES MODEL OF INSTRUCTIONAL PROCESS

Table 124

CLASSROOM OBSERVATION VARIABLES INCLUDED IN EACH INSTRUCTIONAL COMPONENT--COOLEY MODEL

Instructional Process Component	Variable Number	Variable Name
Opportunity	249	Reading, alphabet, language development/ longitudinal
Motivation	398a	All adult praise to children
	403a	Adult praise, behavior
	82	Wide variety of activities, concurrent, negate
Structure and Placement	146	Large group with teacher/reading, negate*
	257	Teacher with one child, academic activities/ longitudinal
	258	Aide with one child, academic activities/ longitudinal
	259	Volunteer with one child, academic activities/ longitudinal
	240	Texts, workbooks/academic activities*
	239	Math or science equipment/academic activities*
	238	Exploratory materials/academic activities*
	260	Children independent, academic activities/ longitudinal
Instructional Events	358a	All child responses
	360a	Child responses, academic
	406a	Adult positive corrective feedback, academic
	407a	Adult negative corrective feedback, academic
	449a	Adult neutral corrective feedback, academic
	412a	Adult feedback, to child responses to adult academic commands, requests, direct questions
	394a	All adult acknowledgment to children
	396a	Adult acknowledgment to children, behavior

-- initial ability--the baseline test score, and
-- the instructional process components.

Regressions were then run with: (1) the initial ability alone, (2) the instructional process variables alone, and (3) the initial ability and instructional process variables together" (p. 320).

"Emrick (1974) diagrammed the relationships between the coefficients of determination (R^2) and the component of variance in the commonality analysis as shown in Figure 42 where R^2_A, R^2_I, and R^2_A represent the coefficients of determination for the three regression runs, respectively. That is, R^2_A is the proportion of criterion variance accounted for by initial ability, R^2_I is the proportion of the criterion variance accounted for by instructional components, and $R^2_{A,I}$ is the proportion of criterion variance accounted for by initial ability and the instructional components. Then,

$R^2_{A,I} - R^2_I$ = the proportion of criterion variance due uniquely to initial ability.

$R^2_{A,I} - R^2_A$ = the proportion of criterion variance due uniquely to the instructional components.

$R^2_A + R^2_I - R^2_{A,I}$ = the proportion of criterion variance shared by initial ability and instructional components.

$1 - R^2_{A,I}$ = the proportion of criterion variance unaccounted for (residual or error)" (p. 320).

The proportion of criterion variance due collectively to instructional components is further partioned into that due to each of the instructional components.

E. Comments: According to the report, "the value of each component was found by summing the values of its process variables" (p. 319). It is difficult to interpret such sums except in a strictly operational manner which is not really interpreting them. For example, a value for the component "Structure and Placement" would be obtained by summing:

variable 146--"the percent of 'child time' that is spent on" a teacher teaching reading to a large group (p. G-9, G-13) with *inter alia*,

variable 258--"the average frequency per CCL [Classroom Check List]" of incidents in which an aide works with one child on "Numbers, math, arithmetic; Reading, alphabet, language development, Social Studies, geography; Science, natural world" (p. G-18).

The first variable measures proportion of time, the second average frequencies. It is unclear what the sum of the two would represent. Similarly, the former measure has to do with a small adult to child ratio, the latter with a high ratio. It is difficult to conceive what sort of structure is reflected by the sum of both scores.

In addition to the difficulty encountered in interpreting such sums, their stability is open to question.

The report compares the results of this commonality analysis with that conducted with data collected the Spring of 1972 from a different sample of children using different (four of eight) Structure and Placement variables. This summary does not report those comparisons.

F. Results:

1. Means and SD's of First Grade MAT Reading Scores (p. 322):

Variables	1973 First Grade	
	Mean	SD
Predictor Variables		
Entering Ability on Tests	29.60	7.00
Motivation	- .09	2.10
Opportunity	.67	.14
Structure and Placement	- 17.50	18.40
Instructional Events	28.20	9.30
Criterion Variables		
Outcome Test Scores	67.18	16.00

2. The coefficients of determination of the regression analyses were:

R^2_A (ability)	.53
R^2_I (instructional dimensions)	.11
$R^2_{A,I}$ (total)	.63

3. Variance attributable to respective sources was:

Source	Percent of Total Criterion Variance
Entering ability (unique)	52
Instructional processes (unique)	10
Shared (ability and process)	1
Error (residual or unexplained)	37

4. Percentages of criterion variance uniquely accounted for by each instructional process variable in the Cooley model were:

Variables	Percentage
Motivation	2
Opportunity	0
Structure and Placement	0
Instructional Events	2

The report suggests these results indicate ". . . the effect of the combined interaction of the process variables in explaining the criterion test scores is greater than the addition of the percentages of the variances of the single process variables" (p. 325).

The sizable difference between (a) the sum of the coefficients of determination when reading and word analysis are regressed separately on each of the instructional components and (b) the coefficient of determination obtained when reading and word analysis are regressed on all four instructional components suggests a high degree of colinearity among the independent variables.

1.5 Robert S. Soar. Final Report. Follow Through Classroom Process
Measurement and Pupil Growth (1970-71). Gainesville, Florida:
College of Education, University of Florida, June 1973. 231 pp.*

This report reviews three years of work by Soar and his colleagues at the University of Florida. During the first two years, the Florida group conducted preliminary studies, including observations in 70 classrooms (ten sponsors, with eight FT and two NFT for each). Results for the first two years were reported in a preliminary report.** The final report and hence this study summary focus mainly on the third year of study, 1970-71.

This summary of Soar's Final Report is divided into four parts:

I. CLASSROOM OBSERVATION STUDY . . . . p. 252
II. RELATION OF CLASSROOM MEASURES TO PUPIL DATA . . . . p. 263
III. STUDY OF TEACHER BEHAVIOR AT THREE POINTS IN TIME . . . . p. 283
IV. THE STABILITY AND RELIABILITY OF OBSERVATION MEASURES . . p. 286

The bulk of the Final Report and consequently this summary deal with the first two topics.

---

*An appendix to Soar's 1973 report was released in June 1976. This 29-page appendix reported the results of multiple range tests performed on classroom process variables identified via factor analyses described in Soar's 1973 report. The purpose of the range test analysis was to identify differences between sponsors in terms of the classroom observation variables. (Sponsors were not explicitly identified in the 1973 report.) The nine sponsors included in the analysis fell into two to four differentiable ranges depending on the variables used in the multiple range tests. However, few of the differences between groupings proved to be statistically significant.

**R. S. Soar. Follow Through Classroom Process Measurement, Research Report, Institute for Development of Human Resources. Gainesville, Florida: University of Florida, 1971.

I. CLASSROOM OBSERVATION STUDY

A. Research Question

In behavioral terms, what are the differences among the FT programs as observed in classrooms? (paraphrased from p. 1)

B. Sample

A total of 289 classrooms, associated with eight sponsors were included in the 1970-71 study. These classrooms were distributed as follows;

Sponsor code		02	03	05	07	08	10	11	13	Subtotals by Type of Classroom
# of sites*		5	5	5	6	4	6	5	2	
# of classrooms by type	K-FT	11	8	9	8	7	8	8	12	71
	K-NFT	3	2	2	2	2	2	1	1	15
	EK-FT	8	4	8	9	5	10	10	-	54
	EF-NFT	2	2	2	2	1	2	2	-	13
	1-FT**	11	9	10	10	10	9	6	8	73
	1-NFT**	3	2	2	2	2	3	2	2	18
	2-FT	7	8	7	4	-	6	4	-	36
	2-NFT	1	2	2	1	-	2	1	-	9
By sponsor:	subtotal FT	37	29	34	31	22	33	28	20	234
	subtotal NFT	9	8	8	7	5	9	6	3	55
										289

"The sample was drawn nationally--from Tampa, Florida, to Spokane, Washington; from Berkeley, California to Burlington, Vermont. Generally each Sponsor was represented by at least two communities at each grade level, with comparison classrooms drawn from the same schools as program classrooms or from nearby schools at similar socio-economic levels." (p. 6). With respect to the first two years of study, "Insofar as possible, settings and programs were selected for observation where SRI was collecting complete data for pupils." (p. 1). Although not explicitly stated, it seems evident that this consideration also served as a criterion for selection of sites in the 1970-71 study.

This study focused entirely on FT cohorts I and II. During the winter of 1971, the entering kindergarten (EK) and entering first (EF) streams of Cohort I were in first and second grades respectively, while the EK and EF streams of CII were in kindergarten and first grades respectively.

---

* Note that the number of sites listed here does not correspond to the number of communities listed in the Soar report (pp. 2-5). In a number of cases, Soar lists separately three or four small communities which according to the FT data codes are all in a single site (for example, within the same county).

** Only the 91 nonentering-first classrooms were included in Soar's secondary analysis, which pooled items across instruments.

C. Variables

a. Independent Measures: FT sponsors, and NFT.

b. Dependent Measures: The Soar study employed four different classroom observation instruments and two rating instruments. This summary describes all six instruments and then summarizes procedures for collecting classroom observation data.

1) Florida Climate and Control System (FLACCS) classroom observation instrument is designed to assess "nonverbal expression of affect in the classroom, physical movement of teacher and pupils, the groupings found in the classroom, and the extent to which individuals or groups were central in classroom activities." (p. 6). "The instrument was made up of three five-minute observation periods, each followed by five minutes of marking...Four instruments were completed during a day in the classroom." (p. 6A). An example of the FLACCS form is given on pp. 7-9 of Soar's Final Report.

2) Teacher Practices Observation Record (TPOR)..."an instrument developed to measure a teacher's practices in relation to John Dewey's Experimentalism. It consists of 62 sign items of teacher behavior....half the items describe behavior which reflects agreement with Experimentalism and would be espoused by John Dewey; the other half..reflect disagreement. In the original procedure, the observer's task was to check those items which occurred during three ten-minute observation periods. The time periods used in this project have been reduced to five minutes in order to parallel FLACCS.

"[TPOR's] major classifications of items for recording behavior are (a) Nature of the Situation, (b) Nature of the Problem, (c) Development of Ideas, (d) Use of Subject Matter, (e) Evaluation of Pupil's Work, (f) Differentiation of Tasks, and (g) Motivation and Control."(p. 10). An example of the TPOR is given in Figure 2, pp. 11-12.

3) Florida Taxonomy of Cognitive Behavior..."The levels into which cognitive activity is divided [in the Florida Taxonomy] are: (1) Memory, (2) Translation, (3) Interpretation, (4) Application, (5) Analysis, (6) Synthesis, (7) Evaluation.

"Although the initial research plan anticipated using the original form of the instrument, it was found difficult to apply to kindergarten-first grade classrooms, and a new version of the instrument was developed. The modification was developed in two stages. First, observers who had visited the classrooms developed items to represent the levels of the Taxonomy from their memory of the classrooms. Then these items were tried out in tape listening, new ones developed as needed, the old ones modified or redefined. When the items stabilized, the form of the instrument was fixed and tape coding was begun." (p. 13) [An example of the modified Florida Taxonomy coding sheet is provided in Figure 3, p. 14.]

"The data of the Cognitive Taxonomy were analyzed [like that from the FLACCS on the TPOR] with each observation period scored zero or one..." (p. 16)

4) Reciprocal Category System. This instrument is a modification of the Flanders Interaction Analysis "by Ober, Wood, and Roberts... The seven teacher categories of the Flanders System have been expanded to nine...Category 10 remains silence and confusion as before."(p. 17). [A description of the categories of this instrument is given in Figure 4, p. 18.]

5) Classroom Description. "In addition to the observation instruments described above, which were completed on the basis of successive five-minute periods of observation, two additional schedules were used. At the beginning of the day, the observers filled out the first part of the classroom description [Figure 5, p. 19] counting adults and children as well as interest centers and other physical aspects of the classroom. After the day's observation had been finished, the observers completed the remainder of this instrument, recording the time pupils spent in various activities such as meals and snacks, or focused learning with or without an adult." (p. 19).

6) Global Rating. "The observer also made a series of global ratings at the end of the day covering broad aspects of classroom process including the response of pupils and adults to the observers." (p. 14) [Figure 6, p. 20].

Data Collection. "The...data set was collected between the second week of January and the middle of March [1971] assuming that this period would be most representative of the year as a whole. Observers had been trained in a course during the fall quarter; the first week of January was spent in refresher training with a day in Follow Through classrooms in Jacksonville, and data collection began the second week. All teams observed then in Philadelphia, since the number of classrooms there was large enough that all nine teams could work at the same time and could meet after each day's observation to discuss questions raised and agree on common procedures...Seven teams of two observers each were used, consisting of graduate students and research assistants, in addition to two part-time teams of central office staff." (p. 19).

"The data collection procedure called for a team to spend a full day in the classroom. Teams arrived at their schools early to meet the teacher before school began (although the delay at the school office sometimes prevented the early meeting with the teacher). As the class was getting under way, observers filled out the first part of the classroom description, and then began the observation. One observed using FLACCS, the other the TPOR. Then, after completion of two instruments, the observers changed instruments and completed two more. Simultaneously with each observation, the observers also made a tape recording, which was later coded on the RCS and Cognitive Taxonomy. During recess or a free-play period, the observers measured the classroom and talked with the teacher so as to permit her to ask any questions she wished about the observation. After the day was over, the remainder of the classroom description and the global ratings were completed." (p. 23).

The dependability of the observation instruments and the reliability of the factors derived from them are described in the summary of the "Teacher Behavior of Three Points in Time" study.

D. Analysis

Analysis of Observational Data

"The first step in the analysis of... observational data was to calculate means and standard deviations for each of the items, for all teachers. Items with very small means and/or variances were then either eliminated or pooled with related items. Following this procedure, an area transformation was carried out item by item to make the data as nearly normally distributed as possible, and with approximately equal variabilities. The first of these considerations seemed important since many items showed essentially "J-curve" distributions; and equality of variances was important because incomplete factor scores were to be calculated later which would involve simply summing and averaging a series of items of behavior without further weighting.

"The data for each observation system were next factor analyzed separately using principle components factor extraction with multiple R squared in the diagonal, followed by varimax rotation of a series of numbers of factors. Rather than relying on one or more of the various rule-of-thumb criteria for number of factors to rotate, a series of factors were rotated, and the output interpreted. The number of factors rotated which seemed to offer the clearest set of interpretations was retained..."(p. 26).

"After the decision had been reached about the number of factors to be rotated for each observation schedule, incomplete factor scores were calculated by simply pooling the T-scores for those measures which loaded ± .40 or above on each of the factors...Following calculation of factor scores, differences between program and grade level means were tested using Duncan's new multiple range test.*" (p. 25). The factor analyses were conducted separately on each of the classroom observation instruments and also on two rating scale instruments together.

Combined Analyses Across All Observation Instruments (pp. 130-168) After preliminary analysis of data from each instrument, a secondary factor analysis was performed on a set of items pooled across instruments. The combined analysis was limited only to data for the 91 non-entering first classrooms. (p. 130)

Included in the combined analyses were "items which loaded most heavily on each factor of each instrument....several analyses were tried in exploratory fashion." (p.130). Specifically, 76 items from the six instruments were included in four separate combined analyses:
- Transpose Factor Analysis of Distances (p. 130)
- Profile Analysis of Items (p. 134)
- Factor Analysis of High Loading Items (p. 134)
- Profile Analysis of Factor Scores (p. 146)

E. Comments

The biggest problem in interpreting the Soar results is that the study contains both too much and at the same time too little information. Having used six different instruments in observing FT and NFT classes, this study generated roughly five hundred items of information for each of the 289 classrooms observed, in other words, roughly 150,000 pieces of information, even though classes were observed only one day each. In order to deal with this massive amount of information, some sort of information sifting or data reduction had to be employed. The process mainly employed was factor anlaysis, but this process also employs abstraction which makes it difficult to relate results to specific behaviors actually observed in classrooms. For example, the profile analysis of factor scores (p. 146) was performed on factor scores of 76 items which had been selected on the basis of a preliminary analysis of roughly 500 items from six instruments, which were selected for unreported reasons.

In addition to the abstraction problem, the study omits certain key pieces of information. For example, when the multiple range test was applied to the factor scores from the original factor analysis, the report notes which factors discriminate homogeneous subsets of teachers associated with different sponsors (p. 114) but the study does not report exactly which sponsors were discriminated by which factors. (The text notes that limited information of this type it provided in Appendix 5 [p. 113], but the report contains no such appendix.)*

---

* W. J. Dixon. BMD Biomedical Computer Programs 2nd Ed. (Los Angeles: Health Sciences Computing Facility, University of California, 1970).

Also it is unfortunate that there is not more continuity between the analyses performed. For example, a multiple range test on the factor scores derived from the factor analysis of high-loading items would have helped to determine whether pooling items across instruments increased or decreased the power of factor scores to discriminate between grade level and sponsors.

Beyond these caveats and those general to factor analysis, this study seems to have been performed carefully and reported modestly.

F. Results

Results are described in two parts corresponding to original analysis of observational data and the combined analysis.

Original Analysis of Observational Data.

The four observation instruments altogether contained "over 400 items and measures of behavior." (p. 51). The two rating instruments contained approximately another 65 items. Data from the four observation instruments were factor analyzed separately while those from the two rating instruments were pooled prior to factor analysis. The factors derived from these analyses and the number of items comprising each, together with the results of multiple range tests on factor scores by grade and sponsor are shown in the following table which has been adapted from Soar's Table 102, pp. 114-15.

"The results as a whole indicate the power of these observational and rating measures to discriminate between groups. A total of 39 factors were analyzed out of the six instruments,...of which 32 discriminated between experimental programs at the one percent level of significance, and one discriminated at the five percent level. Of the remaining six factors, three discriminated significantly between grade levels, leaving three which failed to discrminate in either case--all in the Cognitive Taxonomy. These data are especially relevant since the reliability data on the instruments which are cited later are inadequate in some ways, and these data help to fill the gaps.

"Every factor from FLACCS discriminated significantly...

"From the TPOR, six of the seven factors discriminated between sponsors at the one percent level....

"All nine of the factors from the RCS discriminated between experimental programs at the one percent level. This degree of power from data coded by someone who only knows the classrooms from what he hears on tape seems surprising.

---

*As noted in the footnote on page 251, this appendix was released in 1976.

Soar Table 1: Factors Derived from Classroom Observation and Rating Instruments
(Adapted from Soar, Table 102, p. 114.)

	No. of Items in Factor	Results of Multiple Range Tests			
		Grade		Sponsor	
Factor Description		Homogeneous Subsets	F	Homogeneous Subsets	F
Florida Climate and Control System					
1. Strong Control	22	2	3.10*	2	2.08*
2. Pupil Free Choice vs No Choice	14	2	5.28**	4	28.79**
3. Teacher-Pupil Supportive Behavior	11	1	.25	4	7.34**
4. Nonverbal Gentle Control	4	1	.39	2	3.27**
5. Gentle Control	6	2	3.17*	4	4.95**
6. Work Without Teacher	3	2	9.16**	4	4.60**
7. Pupil Negative Affect	12	1	1.63	3	4.03**
8. Teacher Attention in a Task Setting	5	2	2.38	3	6.09**
9. Teacher Positive Affect	10	2	2.99*	3	2.98**
Teacher Practices Observation Record					
1. Convergent Teaching	7	2	5.81**	5	15.55**
2. Experimental Teaching	13	2	2.97*	5	10.91**
3. Teacher Discourages Exploration	7	2	2.45	3	4.80**
4. Undifferentiated Teaching	3	2	4.50**	3	15.04**
5. Pupil Free Choice vs Teacher Structured Activity	20	2	8.23**	7	28.91**
6. Unnamed	4	2	3.67*	1	1.08
7. Exploration of Ideas vs. Textbook Learning	16	2	3.00*	6	47.65**
Reciprocal Category System					
1. Varied Pupil Initiated Interaction vs. Response to Teacher	18	3	6.95**	4	17.85**
2. Teacher Response and Amplification	8	2	2.38	4	4.69**
3. Drill	12	3	7.24**	4	23.52**
4. Teacher Direction and Criticism vs Teacher Indirect	12	1	1.04	3	6.14**
5. Extended Teacher Talk	10	2	3.24*	3	2.75**
6. Pupil Talk	6	2	2.63	3	2.59**
7. Teacher Acceptance vs Teacher Correction	5	1	.05	3	3.63**
8. Supportive Pupil Talk	6	1	.55	3	3.60**
9. Teacher-Pupil Interaction in Accepting Climate	17	3	5.[illegible]**	3	3.91**
Taxonomy of Cognitive Behavior					
1. Memory	3	1	.87	5	27.31**
2. Applying Previous Learning	5	1	1.57	1	0.80
3. Reading	4	3	32.11	3	7.62**
4. Naming	3	2	3.46*	2	1.48
5. Academic Skills	6	3	3.69*	3	4.53**
6. Unnamed	2	2	3.94**	1	1.31
7. Classification	3	1	.03	2	1.20
8. Information Giving and Receiving	5	1	.25	1	1.17
Global Ratings and Classroom Description Measures					
1. Informal vs Formal Classroom Orgal.	11	2	5.74**	4	25.41**
2. Climate	8	2	2.86*	3	4.43**
3. Structured Learning w/o T. vs. w/T	4	1	.37	4	7.06**
4. Percent Nonwhite	4	3	8.45**	2	6.28
5. Time vs Space	4	3	48.34**	4	5.71**
6. Unstructured vs Structured Time	4	3	3.99**	4	8.56**

[1] N = 289 classrooms   *p < .05   **p < .01

"The Cognitive Taxonomy appears to be the weakest instruments in the battery, so far as power to discriminate between programs is concerned, although its first factor, Memory, showed one of the largest F ratios of any of the factors for discrimination between the programs, and Factor 3, Reading, showed one of the largest F ratios between grade levels...." (p. 113)

"Every factor from the GRCD discriminated programs at the one percent level, and only one factor failed to discriminate between grade levels.

"Taken as a whole, the battery of classroom measures appears to discriminate rather powerfully between both grade levels and experimental programs, which argues both for the reliability of the measures and for their usefulness as program descriptors." (p. 116)

Combined Analyses

As mentioned above, four secondary analyses were performed on the 76 high loading items pooled across the six instruments for the 91 non-entering first classrooms.

Transpose Factor Analysis of Distances

"The first analysis applied as a way of grouping teachers (and programs) on multiple variables was Guertin's distance (d) analysis (Guertin & Bailey, 1970; Guertin, 1971)*, which is a transpose factor analysis of the d statistic. In the results cited to this point in this report, factor analysis has been used as a way to group the observational data into a smaller number of measures which are more reliable and less redundant. What factor analysis, in its usual form, does, is to group measures which are highly related into factors. A parallel statement would be that it puts together into one factor measures which create similar profiles across people. Another variant of factor analysis, called transpose analysis, takes its name from the fact that the roles of people and measures are transposed from those of traditional factor analysis, so that what the analysis does, in effect, is to factor people rather than measures. That is, it groups people who have similar profiles across measures, rather than grouping measures which have similar profiles across people, as the usual factor analysis does." (pp. 130-31).

---

* Guertin, W. H. The search for recurring patterns among individual profiles. *Educational and Psychological Measurement*, 1966, *26* (1), 151, 16

Guertin, W. H., & Bailey, J. P. *Introduction to modern factor analysis*. Ann Arbor, Mich.: Edwards Brothers, 1970.

The transpose analysis resulted in four factors (or groupings) of teachers. "Factor 1 includes the largest number of teachers of any of the factors--indeed a significant proportion of the total number of teachers. Several programs have a large majority of their teachers loading heavily on this factor--Program 1, Program 5, and Program 7 [i.e., Bank Street, Far West, and Arizona]. Factor 2 contains half the teachers from Program 2 [Oregon] and only one or two from most of the other programs. Factor 3 contains five out of six teachers from Program 3 [EDC], three out of ten from Program 1 [Bank Street], three out of nine from Program 6 [Florida], and only a scattering from other programs. Factor 4 contains six out of ten teachers from Program 2 [Oregon], which also was heavily represented on Factor 2, ten out of 18 from Program 9 [NFT comparison teachers] and only a scattering from the other programs. When additional factors were rotated, they consisted of only three or four teachers, again scattered across programs." (p. 131).

Profile Analysis of Items

"Since the groupings from the d analysis appeared rather coarse, and since information about the basis for the grouping appeared to be desirable, Guertin's Profile Analysis (Guertin, 1966; Guertin & Bailey, 1970) was applied to the same set of 76 items. These results consisted frequently of relatively small groups of teachers which were identified as having similar profiles, and the profiles were typically identified by 20 or 30 items on which this group of teachers deviated by more than a standard deviation from the mean of the total group. Interpretation of these restuls was difficult, in the sense that a great deal of information was presented for assimilation. The problem appeared to be one of finding a suitable compromise between the amount of information retained and the difficulty of interpretation." (p. 134).

Factor Analysis of High Loading Items

"As an approach toward reducing the complexity of the profile analysis output, the 76 items drawn from previous analyses were subjected to R analysis. (Another alternative would have been a second-order factor analysis of the factor scores from the five instruments. We have carried out this procedure on several occasions earlier, and in each case it has appeared to combine such large amounts of data that the meaning was essentially destroyed. Factors which had separately discriminated between programs and related to the growth of pupils, when subjected to second order factor analysis produced results which had neither of these validity characteristics.)

"As might be expected, the data from this combined R analysis appeared to be somewhat more complex than that from individual observation instruments, and a larger number of factors (11) were required to represent what appeared to be the clearest structure." (p. 134).

The eleven factors from the high-loading items factor analysis were

1. Teacher-Pupil positive Interaction (8 items)
2. Teacher-Pupil Negative Affect vs. Positive Climate (10 items)
3. Teacher Asks Hard Question, Pupil Mulls (6 items)
4. Gentle Teacher Control and Support (5 items)
5. Pupil Counts, Adds, Interpretation and Memory (5 items)
6. Teacher-Pupil Translation (3 items)
7. Divergent vs. Convergent Teaching (12 items)
8. Teacher Indirect vs. Criticism (6 items)
9. Teacher Central vs. Pupil Selected Activity (13 items)
10. Teacher Uses Text, Prevents Ambiguity (6 items)
11. Structured Activities Without the Teacher (5 items)

Profile Analysis of Factor Scores

"Having reduced the data from 76 items to 11 factor scores, we were ready to reapply the profile analysis procedure....The profile analysis procedure..utilizes both of these analyses [i.e., the normal factor analysis and the transpose factor analysis] in order first to identify profiles which are similar in shape; but then, among those which are similar for shape, to identify those which are similar for shape, to identify those which are similar in level.... The output of the analysis, finally, then, is a series of profiles which are similar first for shape (shape families) and then for level (patterns)." (p. 146).

"This analysis results in twelve 'shape families' of factor score, that is groups of teachers with similar profiles for the eleven factors identified from the factor analysis of high-loading items. After discussing each of the shape families, Soar concludes, "Perhaps the most compelling aspect of these results is the complexity of teacher behavior which is reflected....Programs are no longer lumped together in large groups as was true of the D analysis [i.e., the transpose factor analysis of distances]; rather, they are scattered much more widely across shape families and patterns. But these results do not appear to suggest a very high degree of consistency of classroom behavior within program, in that scanning down the column represented by each program does not indicate strong clustering. Subjectively, the greatest degree of clustering appears to be that of Program 6 [i.e., the Florida Parent Education Model] within Shape Family 2; but even here there are relatively large differences in behavior--especially differences in emphasis on the factors reflecting cognitive level of interaction.

"But, it should be remembered that the measures were not chosen to represent the objectives of the sponsors, but rather to identify differences between teachers. Perhaps it is not surprising that homogeneity of classroom behavior within program is not high when the complexity of the behavior is examined by a relatively fine-grained analysis." (pp. 158 & 165).

II. RELATION OF CLASSROOM MEASURES TO PUPIL DATA

A. Research Question:
What is the relation between behavioral dimensions observed in classrooms to pupil growth on achievement scores and change in attitudes? (paraphrased from p. 1.)

B. Sample:
"From the 289 classrooms observed, pupil data were available from SRI for 169 classrooms. The data were initially screened to eliminate pupils for whom complete pre- and post-achievement data were not available, then were sorted by ethnic group and Follow Through status to obtain the four subgroups of advantaged-disadvantaged, white and nonwhite. A number of additional pupils were lost because these data were incomplete, leaving 150 classrooms for analysis, of which 57 were kindergarten, 20 were entering first, 53 nonentering first, and 20 second grade." (p. 30)

"Attitude data from SRI were available for 169 classrooms. Since preliminary analyses suggested differences between advantaged and disadvantaged and white and nonwhite subgroups, data were screened out which lacked these classifications, leaving 165 classrooms." (p. 50.)

C. Variables:
1. Independent:
The classroom observation variables employed in this study are the same as those developed in the classroom observation study, that is, the 39 factor variables derived from the six instruments employed in that study.

2. Outcome:
In the first two years of the Soar study as well as in the third year, which is the focus of this report, the test battery administered for the FT evaluation was not a standard one. "The third year, SRI administered test batteries consisting of half-length standardized tests, as well as experimental tests assembled from items provided by sponsors to represent their objectives. Kindergarten classes were given the Lee-Clark Reading Readiness Test, two of the subtests from the N.Y.U. Early Childhood Inventory, the Caldwell-Soule Pre-school Inventory, the Wide Range Achievement Test, and a self-concept measure. Entering first grade classes (those without kindergarten experience) were given the same battery, plus the Metropolitan Readiness Test and a book of sponsor items. Nonentering first grade pupils (those with experience in Follow Through kindergartens) were given the Metropolitan Readiness Test, two tests made of items supplied by sponsors, the Wide Range Achievement Test, and the same self-concept measure. Second grade pupils were administered subtests or items from the Metropolitan Readiness and Achievement Tests, the Stanford Achievement Test, the Wide Range Achievement Test, and a measure of test anxiety." (p. 27)

Given this lack of standardized testing, one of the major problems facing the Florida evaluators was the development of usable outcome measures from the potpourri of tests which had been administered. They report: "Various procedures were tried and abandoned, but a few conclusions have been provisionally accepted.

> "1. Pupil data should be analyzed into measures differing in degree of complexity or abstractness, since there is considerable independence of these measures, and since past work suggests that different learning objectives respond best to different teacher behaviors."
>
> "2. Pupil measures should be analyzed using gain measures rather than measures of pupil standing at some point in time."
>
> "3. The shape of the relation between pupil pretest and gain should be studied before other analyses are carried out."
>
> "4. Regressed gain should be calculated separately for subgroups whose pretest means differ (cf, socio-economic status groups, ethnic groups), which is not usually done when analysis of covariance is applied." pp. 27-28.

"Analysis of the subtest scores from the first year's data had indicated that the factor structure of regressed gain scores was appreciably different from that of either pre- or post-scores. It was only in the analysis of regressed gain scores that the simple-complex structure emerged clearly..." (p. 28)

Regressed gain scores were calculated as follows. First raw gain scores were plotted against pretest scores. Then "regressed gain was calculated as deviations from a curve fitted to the data..." (p. 37). In some cases linear (pp. 37 & 41) and in other instances nonlinear curves were used to fit the data. (p. 48.)

"Before items were factor analyzed the second year, those which 'topped out' or had very low variability were eliminated. Despite the problems of measures with varying scale length occurring in the same analysis and the reliability problems cited earlier, reasonably clear structures were obtained from factor analyses of the kindergarten and entering first grade, analyzed separately. In addition to the simple and complex factors found the first year, a third factor emerged which appeared to represent skills-learning such as reading, spelling, and arithmetic. It was apparently associated with the addition of the Wide Range Achievement Test (WRAT), which is heaviliy weighted with these skills." (p. 28.)

"For the nonentering first grade data, items and subtests which were low in variance but appeared to be abstract came together with other low-variance items which appeared to measure simpler kinds of learning, suggesting a factor made up of low-variance items. Inspection of the data also suggested a tendency for items to be grouped on the basis of whether they were individually or group administered. Various combinations of items and subtests were combined and factor analyzed, but no really satisfactory structure emerged. Instead, on the basis of the factor analysis and also a priori judgment, composites of items and subtests were formed to represent group and individually administered, simple-concrete, skill, and complex-abstract measures. An item analysis of all items against these composites was carried out, and items added to the composite accordingly." (p. 29.)

"In the third year's data for kindergarten and continuing first grade where the test battery was unchanged, regressed gain was estimated separately by subgroup for items and for the same composites as the second year, and correlations of item gain with composite gain were calculated to verify item placement in composites...." (p. 29.)

Placement of achievement "items" (in many cases these were really subtest scores) into composites is given in Table 2. At each grade-level, simple-concrete, skill and complex-abstract composites were formed. However, at entering first, nonentering first and second grade levels, some of the composites were divided into individually-administered and group-administered test composites. Specifically, for entering-first, the complex-abstract composite was subdivided in this way, for the second grade level the skill and the complex-abstract composites were both subdivided, and for nonentering first all three types of composites were used in "group" and "individual" forms.

Using these thirteen achievement measures in his third year of inquiry, Soar found that "Study of the regressed gain scores showed widely deviant scores for a few pupils which led to an intensive examination of the regressed gain data. Both ceiling effects and strongly non-linear relations between pretest and gain scores were found when these relations were plotted. As a consequence, the data were reanalyzed, beginning with composite pre- and post-scores. It was generally necessary to drop pupils with high pretest scores in order to lessen the ceiling effect (it could not be eliminated in all cases without doing more violence to the sample than seemed wise). In addition, for a number of composites pronounced nonlinearity of relation between pretest and gain still existed, so a program was developed to fit second degree curves to the data, and calculate regressed gain as the deviation from the curve. As before, each analysis, linear or nonlinear, was carried out separately for four subgroups (socio-economic status, defined by Follow Through status, and white and nonwhite). (p. 29.)

Table 2 : Composition of Soar Outcome Measures by Grade Level*

KINDERGARTEN	ENTERING FIRST	NONENTERING FIRST	SECOND GRADE
Simple Concrete (5) NYU Alphabet NYU Numerals WRAT Naming Letters WRAT Oral Numbers WRAT Recognizing Letters	Simple Concrete (7) NYU Alphabet NYU Numerals WRAT Naming Letters WRAT Counting Dots WRAT Recognizing Letters MRT Alphabet SIG Count & Write Numbers	Simple Concrete-Individual (2) WRAT Naming Letters WRAT Oral Numbers Simple Concrete-Group (3) MRT Alphabet SIG Order of Alphabet SIG Count and Write	Simple Concrete-Individual (2) SIT Days of the Week WRAT Naming 13 Letters
Skill (5) WRAT Spelling from Dictation WRAT Which is More? " Solving Oral Probs. " Written Computation " Word Reading Aloud	Skill (6) WRAT Spelling from Dictation WRAT Solving Probs. " Written Computations " Word Reading Aloud MRT Numbers SIG Order of Alphabet	Skill-Individual (7) SIT Word Reading SIT Numeral Reading SIT Add & Subtract WRAT Spelling from Dictation WRAT Which is More? " Written Computation " Word Reading Aloud Skill-Group (4) MRT Numbers SIG Make Sides Equal " Number-Line Drawing " Adding-Balancing Equs.	Skill-Individual (9) SIT Word & Phrase Read SIT Reading Sounds SIT Story Reading SIT Comprehension SIT Reading Numerals WRAT Spelling " Which is More? " Written Computation " Word Reading Aloud Skill-Group (4) MRT Numbers SAT Word Reading 1-10 SAT Word Reading 11-20 MAT Arithmetic
Complex Abstract (7) PSI Concept Activation - Sensory PSI Concept Activation - Numbers LC Letter Crossout LC Matching Letters & Words NYU Numerals WRAT Counting Dots WRAT Oral Numbers	Complex Abstract-Individual (4) PSI Social Responsiveness PSI Assoc. Vocab. " Concept Activation - Sensory PSI Concept Activation - Numeric.	Complex Abstract-Individual (5) SIT Hidden Figures SIT Verbal Opposites SIT Similarities SIT Absurdities WRAT Solving Oral Problems Complex Abstract-Group (3) NYU Matching Letters & Words MRT Matching MRT Copying	Complex Abstract-Individual (4) SIT Opposites " Verbal Opposites " Similarities " Absurdities Complex Abstract-(Group (2) MRT Word Matching MRT Copying

Source: Table 8, p. 34; Table 9, p. 35.

*Key of acronyms for sources of test items (mode of administration):
NYU - New York University Early Childhood Inventory (Group)
WRAT - Wide Range Achievement (Individual)
PSI - Preschool Inventory (Individual)
LC - Lee-Clark Reading Readiness Test (Group)
MRT - Metropolitan Readiness Test (Group)
SIG - Sponsor Item (Group)
SIT - Sponsor Item (Individual)

Attitude scores were apparently developed in a manner similar to that for the achievement measures. However, the report does not make clear what number of attitude items were available for inclusion in the attitude composite, nor is it clear exactly how some of the items mentioned were incorporated into the composite. "Factor analysis of the [attitude] pretest items produced lower loadings overall than analysis of the achievement items had, suggesting lower reliability for this group of items. Accordingly, composites were created on the basis of the analysis of pretest items for both pre and post data, and the composites then reduced to regressed gain, and classroom means calculated.

As shown in Table 3, "a first factor was obtained which was quite stable across grade level groups. Additional factors were not stable across grade levels, were not often interpretable, and usually were made up of small numbers of items. The first factor, then, was used to represent attitudes toward school at all grade levels." (p. 50.)

The other outcome measure used in this study was number of days absent.

Covariates:
Covariates were used in some of the secondary analyses as explained below.

D. Analysis:
Four different kinds of analyses were performed in this study: correlational analysis (which forms the bulk of the effort), a regression analysis with interaction terms, and an analysis of variance involving the profiles developed in the "classroom observation study."

Correlations:
Following calculation of regressed gain scores, "classroom means were...calculated for the regressed gain measures, T-scored, and related to the classroom process factor scores as well as the rating scales and classroom descriptions." (p. 29.) Specifically, correlations were calculated between each of the various outcome measures, and the 39 factor scores developed form the classroom observation and rating instruments. "In addition, in order to clarify relations further, pupils were separated into subgroups on the basis of socio-economic status and ethnic group and classrooms were divided on the basis of city size, wherever sample sizes permitted, and relationships of growth to classroom process were reanalyzed." (p. 29.)

Table 3

Factor Loadings for Attitude Data

Item	Grade Level			
	Kindergarten	Entering First	Nonentering First	Seco
Put a mark on the face that shows how you feel/think...				
1. when you eat ice cream.			.45	
2. when you fall down on the playground.				
3. about learning out of books.	.58	.44	.47	.43
4. when you think about coming to school in the morning.	.64	.55	.46	.49
5. about the boys and girls in school.	.56	.41	.52	.42
6. about learning new things.	.71	.66	.52	.40
7. the boys and girls in this class feel about you.	.66	.47	.52	.47
8. the teacher feels about you.	.70	.49	.57	.44
9. feel about your teacher.	.67	.56	.61	.49
In this box is your teacher, and in this box are your friends,				
10. put a mark where you want to be.				
Here is your teacher and here are five empty circles,				
11. put a mark on the empty circle where you want to be.				
Here are your friends and here are five empty circles,				
12. put a mark on the empty circle where you want to be.				

ource: Soar, Final Report, p. 50a.

Regression using interaction terms:

The bulk of the Soar analyses relating classroom measures to pupil data were the zero-order correlations just described. However, initial results made it apparent that "a number of measures of classroom behavior either failed to relate in the expected direction with pupil gain, or related in the opposite direction" (p. 192). Speculating that "one of the possible explanations for these unexpected and contradictory results could be the interaction of other variables" (pp. 193-194), Soar examined this possibility by "using stepwise multiple regression, entering the product term (which tests for interaction) following the entry of the two measures themselves as main effects." (p. 194.) "The sample [for this study] was nonentering first grade, since that was one of the two largest, and represents other grade levels better than kindergarten. Mean classroom regressed gain for the Group Skill and Group Abstract measures were used as dependent variables. The process measures tested were: FLACCS 2 X FLACCS 9, FLACCS 5 X FLACCS 9, FLACCS 6 X FLACCS 9, TPOR 1 X FLACCS 9, TPOR 2 X TPOR 5, TPOR 5 X FLACCS 9, TPOR 5 X RCS 8, RCS 8 X FLACCS 9, Cog Tax 3 X FLACCS 2, Cog Tax 3 X FLACCS 9, Cog Tax 8 X FLACCS 2" (p. 194.)

Differences in Pupil Gain Associated With Profiles

In a third kind of analysis Soar attempted "to test whether pupils experiencing the different patterns of teaching identified by the profile analysis gained in achievement at differing rates. To test this, the nonentering first grade classrooms which made up each profile were identified, and the mean gain for pupils on each of the measures was entered into analysis. The classroom was used as the unit of analysis. Where a classroom was included in more than one profile, it was placed with that profile for which its "d" was smallest, and only entered there.

"Only 43 classrooms were available which were observed as nonentering first grade and for which pupil achievement data were available. One of the profiles included six teachers, five profiles included three, six profiles included two teachers, and ten teachers were not in any profile." (p. 201.) Differences in gain scores associated with the different teacher profiles were tested with a "single factor analysis of variance." (p. 201.)

E. Comments:

The bulk of this study consisted of the correlation analyses described above. Given this it is unclear why Soar switched to regression analysis to investigate the interaction of the classroom observation measures as they relate to outcome measures. It would have been more in keeping with the initial approach, if Soar had used partial correlation to investigate the multivariate case.

There are, of course, some drawbacks to the use of correlational analysis in the first case. Specifically, the correlation approach contributes to a degree of abstraction which makes it difficult to relate findings back to any of the behaviors observed in classrooms. Findings are expressed mainly in the form classroom observation factor X correlates significantly with outcome factor Y. Since the factor scores themselves are developed from standardized item and subtest scores, it is virtually impossible to know even what ranges of classroom behaviors or of test scores might be involved. This abstraction problem is, of course, not really the responsibility of the Florida evaluators, since they obviously had to employ some form of data reduction in order to make sense out of the variety of outcome measures available. Nevertheless, the abstractness of the analysis does make it difficult for the reader to interpret findings.

In most respects Soar's findings are interpreted in the final report with restraint and attention to limitations of the analysis. For example, he notes "The problem of an adequate gain score is made more severe than it would be otherwise by the fact that Individual Skill is a considerably longer subtest than any of the others....It seems likely that sponsors who stress more complex objectives would be penalized by [the] procedures [used in this analysis]." (p. 48.)

"The problems that were encountered in calculating pupil gain have been exceedingly difficult...and the limitations of these gain measures must also qualify the conclusions drawn from subsequent analyses in which these data are used as criterion measures. We cannot be sure what classroom behaviors promote the growth of pupils if we cannot be sure which pupils grow more than others." (p. 50.)

In two respects, however, more conservative alternative interpretations than those offered by Soar might be offered.

1. Since only eight percent of the correlations between the attitude measure and the classroom process measures (across all four grade level streams) were statistically significant at the .05 level and only one out of 154 was significant at the .01 level, a conservative interpretation might have attributed these findings to mere chance, instead of offering substantive interpretations for them.
2. Since not even one of the classroom observation variables consistently correlated at a statistically significant level across grade levels with any of the achievement outcome measures, a conservative interpretation might attribute differences in correlations across grade levels to chance variation, rather than to substantive processes.

Another alternative conservative interpretation can be offered with regard to Soar's correlation analysis. This interpretation is more easily explained after presenting Soar's results; so it is postponed until the end of the next section.

F. Results: Relation of Classroom Measures to Pupil Data

"In beginning to discuss the data relating classroom measures to the measures of pupil gain, several comments seem appropriate. The correlations for nonentering first grade data are generally higher than those for kindergarten (the two larger groups), and the correlations for Individual Skill tend to be among the highest of those at nonentering first grade. The number of cases for entering first and second grades were much smaller. All of the results which follow should be interpreted in the light of the difficulties found for the pupil gain measures. A ceiling effect was found in numbers of the pupil measures at kindergarten and nonentering first grade which was often accompanied by a nonlinear relation between pretest and raw gain, so that the gain of pupils with high pretest scores was sharply restricted, both with respect to the mean gain which was possible for them, and with respect to the variability which occurred around the mean. Regressed gain was calculated by fitting curves to numbers of the measure and calculating gain as variability around the curve. High scoring pupils were also eliminated from the analysis, but the problem of restricted variability above and below the curve for high prescoring pupils remained for a number of measures. It seems possible that the results which follow may reflect to an unknown degree the inability of higher standing pupils to show an appropriate amount of gain because of measurement difficulties for these two grade levels." (p. 168)

"In general, these results for the relations between classroom behavior and pupil gain differ from previous work in numbers of ways. The current results suggest that classroom measures such as highly structured teaching, with a relatively narrow subject matter focus and little freedom for pupils were associated with pupil gain. As one approach toward identifying the source of these differences in results, pupils were broken up into subgroups differing by socioeconomic status and ethnic group, but the number of classrooms was found to be too small for analysis for white low and nonwhite high social status groups. When additional pupils were dropped because of the ceiling effect, the white high subgroup also had too few classrooms for analysis.

"Another approach to the problem was attempted--that of dividing the the data by city size. Analyses of classroom behavior with mean pupil gain were carried out in these two classifications, with the findings that the correlations frequently differed from large to small cities. However, the distribution of sponsors differed materially with city size, and in different ways from kindergarten to first grade. If the relations between classroom behavior and pupil gain were carried out in these two classifications, with the finding that the correlations frequently differed from large to small cities. However, the distribution of sponsors differed materially with city size, and in different ways from kindergarten to first grade. If the relations between classroom behavior and pupil growth had been parallel across city size groups from kindergarten to first grade, then there would have been justification for interpreting the differences in terms of city size." (p. 168)

"Unfortunately, the smaller number of significant correlations for kindergarten and the differences between kindergarten and first grade for the total group made it difficult to know whether the results were parallel or not. There was also the difficulty that the number of classrooms in small cities was small (N = 14 to 16), enough that those data are highly variable..." (pp. 168-169).

Florida Climate and Control System
"None of the [FLACCS] factors related in consistent manner across the four grade levels...In general, the relations between teacher expression of affect and gain which are generally expected failed to appear. Rather, the dimensions reflecting structuring of classroom activities and pupil involvement with the teacher were the ones which tended to relate to pupil gain. Perhaps the major differences among Follow Through classrooms are not on affective dimensions but rather ones that reflect structuring." (p. 169)

Teacher Practices Observation Record
"These results [for the TPOR] seem to parallel those from the Florida Climate and Control System in indicating a positive relation between narrowly focused management of subject matter or limited pupil freedom, and gain. Again, as with FLACCS, entering first grade sometimes differs but apparently does so because pupils had less freedom there, on the average, than in the other three groups. (p. 178).

Reciprocal Category System
"The negative association between Varied Pupil Initiated Interaction and gain, and the positive association between Drill and gain seem to parallel again the tentative conclusions suggested by FLACCS and the TPOR that a narrowly focused classroom and less pupil freedom are related to subject matter growth. The fact that this instrument was coded by a different set of observers than the previous two "live" instruments--coders who had not seen these classrooms--lends added support to the agreement." (p. 178)

Taxonomy of Cognitive Behavior
"Across the four grade levels as a whole... there were several factors which appeared to relate in reasonable fashion.... Although not completely consistent, there does seem to be evidence here for more complex activities relating to more complex growth measures, and for the activity which relates to growth changing from kindergarten to second grade. The data suggest that some teachers were continuing lower level activities beyond the point where they were functional for growth." (p. 183)

Global Ratings and Classroom Description

"Several factors from the GRCD showed reasonable trends across the four grade levels....The agreement of these several grends with data reported for the previous instruments is considerable, with numbers of relations suggesting with association of structured activity (and time with the teacher) and pupil cognitive growth. (pp. 183 & 192.)

Relation of Days Absent

"In summary, there was no clear evidence of fewer days absent being related to gain, which seems surprising." (p. 192.)

Relations between Observation Measures and Change in Pupil Attitude

"[T]here does seem to be a trend in the data [showing the relations between the observation measures and change in pupil attitude toward school]: The number of significant correlations for entering first grade, and the consistency of the meaning of their direction. The meaning appears to be that these pupils are happiest or like school best when the task is simple, well structured, and under the direction of the teacher....

"Relationships within the other grade levels are scattered, but by and large reasonable. More favorable attitudes in kindergarten, where subject matter demands are presumably less, in general, are associated with informal classroom organization, structured working without the teacher, and a shorter school day (or a more spacious classroom)....

"Attitudes of nonentering first graders were positively related to attention from the teacher, teacher acceptance, teacher structured activity rather than free choice, and textbook learning rather than exploring of ideas. There appears to be a degree of parallel with the entering first grade relationships, but the effects seem not to be as strong....

"The results overall suggest that pupils hold more favorable attitudes toward school in a simple, less demanding environment, if either they are meeting school for the first time, or increased subject-matter demands are made, and the effect is strongest when both experiences are met at once. But by second grade these influences appear to have lost their negative effect." (pp. 198 & 201.)

Results from the basic correlational analysis are presented in forty tables (Tables 212-160 & 162, pp. 160-200). It is from these tables that the foregoing conclusions quoted from the Soar report are inferred. Information from these tables is summarized in Table 4. Altogether 1014 correlations were calculated. Of these 137 or 14% reached statistical significance at the .05 level. The proportion of correlations between the 39 classroom measures and the absence and attitude outcome measures which reached statistical significance was about what would be expected by chance alone (.05, and .08 respectively). A considerably higher proportion (.17) of correlations between classroom measures and achievement

Table 4 : Significant Correlations Between Outcome and Process Measures, by Grade-level Stream *

	Achievement Measured**					
	Simple - Concrete K EF 1 1 2 (I) (G)	Skill K EF 1 1 2 2 (I) (G) (I) (G)	Complex-Abstract K EF EF 1 1 2 2 (I) (G) (I) (G) (I) (G)	Summary: # of SS correlations W/18 achiev. measures across 4 grade stream	Absence K EF 1 2	Attitude K EF 1 2
PLACTS (9 factors)						
1. Strong Control	[illegible]			1	[illegible]	
2. Pupil Free Choice vs No Choice		[illegible]	[illegible]	6		
3. Teacher-Pupil Supportive Behavior		[illegible]	[illegible]	2		
4. Nonverbal Gentle Control	[illegible]			2		
5. Gentle Control				0	[illegible]	
6. Work W/O Teacher	[illegible]		[illegible]	4		
7. Pupil Neg. Affect		[illegible]		4		
8. Teacher Attention in a Task Setting		[illegible]		2		[illegible]
9. Teacher Positive Affect	[illegible]		[illegible]	3 (1 pos, 2 neg)	[illegible]	
TPOR (7 factors)						
1. Convergent Teaching		[illegible]	[illegible]	7		[illegible]
2. Experimental Teaching	[illegible]	[illegible]	[illegible]	5 (1 pos, 2 neg)		
3. Teacher Discourages Exploration	[illegible]	[illegible]		2		[illegible]
4. Undifferentiated Teaching	[illegible]		[illegible]	2		
5. Pupil Free Choice vs Teacher Structured Activity		[illegible]	[illegible]	5		[illegible]
6. Unnamed				0		
7. Exploration of Ideas vs Textbook Learning		[illegible]		3		[illegible]
Reciprocal Category System (9 factors)						
1. Varied Pupil Initiated Interaction vs Response to T.		[illegible]		2		[illegible]
2. Teacher Response & Amplification		[illegible]		1		
3. Drill		[illegible]	[illegible]	5		
4. Teacher Direction & Criticism vs T. Indirect		[illegible]		2		
5. Extended T. Talk	[illegible]	[illegible]		3 (1 pos, 2 neg)		[illegible]
6. Pupil Talk			[illegible]	1		
7. T. Acceptance vs T. Correction				0		
8. Supportive Pupil Talk		[illegible]		1		
9. T/Pupil Interaction in Accepting Climate				0		
Cognitive Taxonomy (8 factors)						
1. Memory	[illegible]	[illegible]		2 (1 pos, 1 neg)	[illegible]	[illegible]
2. Applying Previous Learning	[illegible]	[illegible]		2		[illegible]
3. Reading		[illegible]	[illegible]	5		
4. Naming			[illegible]	1		
5. Academic Skills	[illegible]	[illegible]	[illegible]	4 (3 pos, 1 neg)		
6. Unnamed			[illegible]	1		
7. Classification				0		
8. Information Giving and Receiving		[illegible]	[illegible]	3		

Continued on next page.

Table 4 (contd.)

Global Rating & Classroom Description (6 factors)																									
1. Informal vs. Form. Classroom Org.				-				--	--		-				-				5		-			+		
2. Climate						+				+		+						+	4							
3. Structured Learn. W/O T. vs W/Teach.	-					--		--	-	-	--	--			-			-	9					++		
4. Percent Nonwhite			++	+									-		+			-	5 (3 pos, 2 neg)			+				
5. Time vs Space		--		+	-	++							-						5 (2 pos, 3 neg)					-		
6. Unstructured vs Structured Time						--		--	--	-	--				--	-		-	8							
No. of SS correlations (.05 CL)	2	2	6	8	2	12	3	17	14	5	8	8	4	0	11	7	4	5	117	2	1	4	1	3	6	3
Percentage of SS correlations	05	05	13	21	05	31	08	44	36	13	21	21	10	0	28	18	10	13	17	05	03	10	03	08	16	08
No. of Classrooms	49	20	47	50	20	54	20	50	50	20	20	54	20	20	50	50	20	20		54	20	50	20	59	21	60

* Minus signs are used to show negative correlations between the respective outcome and process measures, plus signs to indicate positive correlations. Single signs indicate that correlations were significant at the .05 level, while double signs indicate significance at the .01 level. Blanks indicate that correlations did not reach the level of significance. Information in this table is drawn from Soar's Tables 121-160, & 162, pp. 160-200.

** Achievement measures are listed by type (that is, simple-concrete, skill, or complex-abstract), by grade level (i.e., kindergarten, entering first, nonentering first, or second) and where appropriate by the mode of administration for the "items" comprising each (i.e., individually (I) or group (G) administered).

measures were significant at the .05 level. The strongest associations in this realm were between the skill outcome measures and classroom measures reflecting convergent teaching (TPOR factor #1) and teacher directed, structured classroom setting (GRCD factors #2 and #6). Only for these three classroom factor variables (out of the total 39 developed in the classroom observation study) did as many as one-third of the correlations with the 18 achievement outcome measures (across all four grade levels) reach significance at the .05 level.

Interactions of Process Measures with Gain

"It will be recalled that a number of measures of classroom behavior failed to relate in the expected direction with pupil gain, or related in the opposite direction....explanations for these unexpected and contradictory results could be the interaction of other variables."(pp. 193-94).

"In order to examine such possibilities, a number of process variables which appeared to reflect positive affect, structuring, controlling, or subject matter focus were tested for the existence of an interaction between the variables, taken two at a time. The analyses were carried out using stepwise multiple regression, entering the product term (which tests for interaction) following the entry of the two measures themselves as main effects...The sample was non-entering first grade, since that was one of the two largest, and represents other grade levels better than kindergarten. Mean classroom regressed gain for the Group Skill and Group Abstract measures were used as dependent variables. The process measures tested were: FLACCS 2 X FLACCS 9, FLACCS 5 X FLACCS 9, FLACCS 6 X FLACCS 9, TPOR 1 X FLACCS 9 TPOR 2 X TPOR 5, TPOR 5 X FLACCS 9, TPOR 5 X RCS 8, RCS 8 X FLACCS 9, Cog Tax 3 X FLACCS 2, Cog Tax 3 X FLACCS 9, Cog Tax 8 X FLACCS 2.

"Most of the interactions were not significant, but two reached significance at the five percent level and two at the one percent level....

"The interaction of TPOR 2, Experimental Teaching, and TPOR 5, Pupil Free Choice vs Teacher Structured Activity [with the group-skill gain measure] is...significant at the one percent level, and accounts for 14 percent of the variance in gain in addition to that for the variables singly." (p. 194.)

"[T]he interaction of FLACCS 9, Teacher Positive Affect, with Cog Tax 3, Reading, as they relate to gain in Group Skill,...is significant at the five percent level, and accounts for seven percent of the variance in the gain variable.

"The interaction of TPOR 2, Experimental Teaching, and TPOR 5, Pupil Free Choice vs Teacher Structured Activities...with Group Abstract gain as the dependent variable...is significant at the five percent level, and account for 11 percent of the variance in gain." (p. 196).

"...the interaction of FLACCS 2, Pupil Free Choice vs. No Choice with Cog Tax 3, Reading, as they relate to gain in Group Abstract.... accounts for four percent of the variance in gain, and is significant at the five percent level....

"The interpretation of these results should be tempered by the fact that they represent four findings out of 24 tests run, so that one result significant at the five percent level would be expected by chance. But the fact that four significant results were obtained, two of which reached the one percent level, suggests that real effects may be present. And the results do appear to follow the crude expectation that led to the tests, namely, that a measure of 'freedom' or positive effect might be associated with gain, if simultaneously there were evidence of focus or structure, or task orientation. And amounts of variance large enough to be of practical importance were identified in some of the tests. All in all, there appears to be support here for the usefulness of examining more than one measure of classroom process at the same time." (pp. 196-98.)

Association Between Pupil Gain and Teacher Profiles

"Although the means [of classroom regressed gains] often differed from each other by 20 points or more on the T-score scale, the differences were not significant in any case when tested by single factor analysis of variance. "While this is not evidence that some patterns of teaching are associated with greater gain than others, when such differences in means are not significant, the results do not minimize the usefulness of the analysis, either. Profile analysis appears to be a promising tool for further application in classroom observation: 1) in relating profiles to differences in gain, given a larger number of classrooms; and 2) in testing the degree to which a sponsor was successful in creating the pattern of classroom behavior be identified as desirable." (p. 201).

Summary of Relations of Classroom Measures to Pupil Growth

"These data relating pupil growth to classroom process differ in important ways from previous findings. Where past results showed greater amounts of freedom to be related to greater amounts of pupil growth, in general, these data show the opposite (although the interactions do qualify that conclusion). Where the Group Abstract measure related more strongly to classroom process in the past, the skill measure was the one which was more frequently and more strongly related in these data, and Concrete and Abstract related less often. Whereas measures reflecting freedom were more likely to relate to pupil abstract growth and only a measure reflecting very narrow structuring related to concrete growth in the previous data, there was no discernible trend for such a differentiation in these data. (These comparison apply to data collected in 1968-69 vs 1970-71 data, since the 1969-70 pupil data were too limited to lead to very clear conclusions)." (p. 201).

"When results differ from study to study as these have, one likely possibility is the presence of a significant interacting variable which is not being recognized. It seems probable that the sample from which these data were obtained differs in important ways from previous samples, but at this point the nature of the differences is not known. Little work has been published in which systematic observation measures have been related to measures of growth of disadvantaged pupils. The earlier data from this project agreed with the larger body of work in which advantaged pupils, often from affluent suburban schools, were studied. These results do not. Rather, they agree with the expectations of some theorists that disadvantaged pupils should need larger amounts of structure or control to maximize learning than would be true for middle or upper class pupils.

"When the interactions of number of the classroom observation measures reflecting expression of positive affect or structuring behavior were tested, few were signficant. The significant ones gave limited support to the idea that pupil freedom or teacher positive affect on one dimension, associated with structure or task orientation on the other, was accompanied by increased gain; but even this findings was not without question.

"Another possible reason for the differences in results from the previous data to these may be the loss of high prescoring pupils in the current data and the limited gain possible for others. Beyond this, regressed gain was calculated separately for the four subgroups of nonwhite and white, qualified and not qualified for Follow Through Services. Because of the lack of sufficient data to permit it, this refinement of analysis was not possible in the previous data. Rather, regressed gain was calculated for the total group (which would parallel the typical covariance analysis). It seems possible, then, that the results reported here may be the "right" results and others in error, even though they agreed with the larger body of findings from advantaged pupils.

"Few relationships were entirely consistent across all grade levels, but some trends appeared to agree across instruments. Probably selecting, integrating, and summarizing is always subjective to a a degree." (p. 202.)

"Given these uncertainties, the major trend of relationships between classroom behavior and pupil growth for this sample of pupils were as follows:

"1. Probably the most consistent findings was a relation between a number of factors from the five instruments indicating that greater amounts of teacher control, structure, focus, and convergence, or lesser amounts of pupil freedom, exploration of ideas, or experimental teaching led to increased pupil cognitive growth, especially in the skill measures.

"2. Factors reflecting both positive and negative affect expression tended to relate negatively to pupil gain, but greater amounts of either tended to occur in less structured settings so that it seemed possible that structure rather than affect might be the influential variable in these results. Tests of interactions gave only limited support for this interpretation.

"3. There was some evidence of the nature of activities which related positively to gain changing across grade level. Lower cognitive level activities which related positively at the lower grades either did not relate or related negatively at higher grade levels. This suggested that teachers carried simple activities past the point at which they were functional.

"4. Work in interaction with the teacher tended to relate positively with gain, whereas work without the teacher or an adult--independent work--tended to relate negatively.

"5. One of the strongest relations in the study suggested that a long school day for entering first graders is negatively associated with gain. [An "hours in school day" item is included in the GRCD Factor #5--Time vs Space which correlates strongly with some of the achievement outcomes for entering first (1) classrooms. See Table 3.]

"6. Relations between the observation measures and change in attitude toward school were not consistent across grade levels but highly consistent within grades. The differences suggested that a 'simple' environment related to positive attitude change for pupils first meeting school or specific subject matter demands.

"The existence of changes across grade levels which seemed reasonable may mean that other "real" changes may have been ignored as inconsistencies. The only long run answer is replication." (p. 203.)

Soar closes with a "final word" regarding the limitations of his study and qualifications regarding his findings. Specifically he notes:

"We have no information here about the effect of differing teaching styles on noncognitive objectives such as self-concept, personality, or responsible independence. Nor is there information about whether these results would apply to younger or older pupils, and there are suggestions that some relationships do differ across the grade levels studied here.

"Further, the information we have about growth in cognitive objectives is clearest for skill growth, only applies to a single school year, and leaves open the question of longer-term relationships." (p. 204).

* * * * * * * * *

An alternative interpretation of the correlation results could be offered on the basis of data provided in Appendix D of the Final Report. As previously mentioned, the main analysis in this study consisted of correlating classroom process factor scores with classroom average regressed gain scores. The results of this analysis are presented in the last section. In addition to this analysis--which is the focus of the text of the report--Soar also repeated this correlation analysis after dividng the data by city size ("small city" and "large city"). This reanalysis was carried out only for the kindergarten and the non-entering first grade levels. Soar reports this reanalysis as follows:

"Unfortunately, the smaller number of significant correlations for kindergarten and the differences between kindergarten and first grade for the total group made it difficult to know whether the results were parallel or not. There was also the difficulty that the number of classrooms in small cities was small (N = 14 to 16) enough that those data are highly variable." (pp. 168-69.)

A more conservative position might argue that the reanalysis including stratification by city size is more valid than the original analysis because pooling across city-size confounds the classroom behavior-regressed gain relationship with effects of city size. Such a conservative interpretation might ask which of the classroom process correlations with achievement outcome reaches the level of statistical significance within both the small city and large city sample as well as in the pooled analysis. The results of such a conservative approach are shown in Table 5. Note that only 7 out of 297 process-achievement correlations and zero out of 66 process-absence correlations are statistically significant at the .05 level (and in the same direction) for the small city, the large city, and pooled sample of classrooms.

While this conservative interpretation indicates considerably fewer significant correlations between process and achievement measures than in the pooled correlation analysis (see Table 4 above), the overall pattern is much the same. That is:

--the process measures correlates most strongly with the skill measure;

--the pattern of correlations indicates that greater amounts of teacher control, structure and convergence or lesser amounts of pupil freedom and work without the teacher is associated with increased gains, especially in the skill measure at the first grade level.

Table 5: Correlations Between Outcome and Process Measures, by Grade-level Stream Which Are Significant (.05) in Both Small-City and Large-City Samples As Well As in Pooled Analysis*

	Achievement Measures										
	Simple - Concrete			Skill			Complex-Abstract			Days Absent	
	K	1 (I)	1 (G)	K	1 (I)	1 (G)	K	1 (I)	1 (G)	K	1
FLACCS (9 factors)											
1. Strong Control											
2. Pupil Free Choice vs No Choice					-						
3. Teacher-Pupil Supportive Behavior											
4. Nonverbal Gentle Control											
5. Gentle Control											
6. Work W/O Teacher			-					-			
7. Pupil Neg. Affect											
8. Teacher Attention in a Task Setting											
9. Teacher Positive Affect			-								
TPOR (7 factors)											
1. Convergent Teaching					+						
2. Experimental Teaching											
3. Teacher Discourages Exploration											
4. Undifferentiated Teaching											
5. Pupil Free Choice vs Teacher Structured Activity					-						
6. Unnamed											
7. Exploration of Ideas vs Textbook Learning					-						
Reciprocal Category System (9 factors)											
1. Varied Pupil Initiated Interaction vs Response to T.											
2. Teacher Response & Amplification											
3. Drill											
4. Teacher Direction & Criticism vs T. Indirect											
5. Extended T. Talk											
6. Pupil Talk											
7. T. Acceptance vs T. Correction											
8. Supportive Pupil Talk											
9. T/Pupil Interaction in Accepting Climate											

Continued on next page.

Table 5 (contd.)

Cognitive Taxonomy (8 factors)											
1. Memory		?			–						
2. Applying Previous Learning											
3. Reading				+			+				
4. Naming											
5. Academic Skills											
6. Unnamed											
7. Classification											
8. Information Giving and Receiving											
No. of smallcity csrms	12	16	16	12	16	16	14	16	16	14	16
No. of largecity csmrs	37	34	34	40	34	34	40	34	34	40	34

---

*Achievement measures are listed by type (i.e., simple-concrete, skill, and complex-abstract), by grade-level (i.e., kindergarten, or nonentering first grade) and, where appropriate, mode of administration (i.e., group or individual).

Minus signs indicate correlations between respective outcome and process measures which were consistently negative and statistically significant (at the .05 level) in the small-city sample), the large-city sample and the pooled sample. Plus signs indicate correlations which were consistently positive and statistically significant at the .05 level.

The question mark indicates the one case in which a correlation was statistically significant and negative in the pooled and the small-city sample but satistically significant and postive in the large-city sample.

Source: Tables 121-160-, pp. 160-193; and Appendix D, Tables 170-185, pp. 220-231.

III. STUDY OF TEACHER BEHAVIOR AT THREE POINTS IN TIME

A. Research Question:

The study of teacher behavior at three points in time was conducted "to examine in a preliminary way, the development of classroom organization for teachers who differed materially in control style." (p. xx)

B. Sample

"...two subsets of teachers were selected from the previous winter's sample to be observed at the opening of school, late fall, and during the winter. These subsets were selected as high and low control teachers on the basis of the FLACCS factor that most nearly represented strong control. Ten high and ten low control teachers were identified initially, but one teacher was lost from one group and two from the other as a consequence of a teacher strike. Three additional teachers were lost because of scheduling difficulties (teacher illness the day of the observation, etc.) after the observers were out in the field. The lost teachers were replaced the day scheduled but, of course, no previous data were available for these three. These teachers were observed the first week of school, again in late October, then again as part of the winter sample." (p. 116)

Although the point is not entirely clear, it appears that only seventeen teachers were included in this study -- that is the original 20 selected minus the three lost because of a teacher strike, without the replacements. (see p. 19)

C. Variables

The same six instruments were used to observe the high and low control teachers at each of the three time points as was used in the winter classroom observation study, that is:

1. Florida Climate and Control System (FLACCS)
2. Teacher Practices Observation Record (TPOR)
3. Florida Taxonomy of Cognitive Behavior
4. Reciprocal Category System
5. Classroom Description
6. Global Ratings

"The fall data were normed from the T-score distributions for the total winter group. A first analysis of the data was done omitting these replacement teachers, with the finding that differences on FLACCS 1 approached significance, but did not reach it. Since a central interest was examination of differences associated with differences in control techniques, the teachers were reclassified, pooling fall data on control with previous year data on control, giving the three replacement teachers a mean rank for the previous year. Analysis of all measures was then carried out." (p. 116)

These measures were factor analyzed, as in the main classroom observation study, that is, each of the first four instruments factor analyzed separately and then the classroom description and global ratings information being pooled prior to factor analysis.

D. Analysis

"'The data from the fall observations were reduced by the same factor scoring procedures as the winter sample. Then the factors were analyzed by a two-factor analysis of variance with repeated measures in which differences between high and low control teacher groups were one factor, and the points in time the other." (p. 117).

E. Comments

Results of this study are worth very little because:

-- The study does not make clear exactly how the "high control" and "low control" teachers were selected.

-- Selection of a sample on the basis of extreme pretest scores is a hazardous endeavor, and given "unreliability" or "nongeneralizability" of the scores on which groups were selected, any longitudinal study will be subject to what is commonly called regression to the mean. Indeed this appears to be precisely what happened in this study at least with respect to some variables. For example, with respect to teacher control, this study concludes "high control teachers started with little pupil freedom and increased it, whereas low control teachers started with high pupil freedom and decreased it." (p. 120)

-- The study does not fully describe the implications of the attrition of three of the 20 teachers originally selected (see above).

-- Winter observations for six of the 20 classrooms in the study may have been biased by unusual conditions (pp. 117, 120 and below).

F. Results

Table 103 (p. 118) summarizes the results of the repeated measures analysis of variance. The study summarizes this analysis as follows:

"In summary, low control teachers exercised less strong control in comparison with high control teachers; they also had more pupil work without the teacher and expressed more positive affect. They did less convergent teaching, differentiated more, and encouraged greater exploration of ideas. They directed and criticized less and showed greater acceptance. All of these appear to be reasonable differences between groups of teachers selected on the basis of differences in the coerciveness of the control methods they used.

"With respect to changes over time, strong control by the teacher, expression of pupil negative affect, application of previous learning (perhaps reviewing), and pupil unstructured time decreased, and teacher acceptance increased. Several other measures were different on the second observation from the first and third; work without the teacher and pupil free choice were higher at the second observation; along with a trend for greater exploration of ideas which was not significant. Undifferentiated teaching was also lower at this point, meaning that differentiation was higher as work without the teacher and pupil free choice were higher. The possibility was suggested that a number of the teachers in the subsample were observed the second week after Christmas, so that the similarity of first and third observations may represent a recycling or beginning a new phase of work.

"There were two significant interactions: one was not interpreted since the factor was unnamed; the other suggested that high control teachers started with little pupil freedom and increased it, whereas low control teachers started with high pupil freedom and decreased it." (p. 120)

IV. THE STABILITY AND RELIABILITY OF OBSERVATION MEASURES

A. Research Questions

-- What is the stability over time of observed teacher behavior? (paraphrased from p. 121)

-- What is the "reliability of observers" used in this study? (paraphrased from p. 121)

B. Sample

The sample used to examine the first question was the same as that in the "study of Teacher Behavior at Three Points in Time," that is the seventeen teachers observed served at three points, start of school, late fall, and winter.

The reliability of observers question was addressed separately for the group of four classroom observation instruments (that is the FLACCS, TPOR, RCS and the Cognitive Taxonomy) and for the two rating instruments (the Global Ratings and the Classroom Description). For the former, no description of the sample for the reliability study is provided. For assessing the reliability of the rating instruments, the sample of 289 observed classrooms used in the Classroom Observation Study was employed.

C. Variables

The variables whose stability was assessed were 37 of the factors derived from the five separate factor analyses reported in the analysis of observation data, that is:

FLACCS -- 9 factors
TPOR -- 7 factors
RCS -- 9 factors
Cognitive Taxonomy -- 8 factors
GRCD (Global ratings and classroom description together -- 4 factors
("Two factors which represented such stable aspects of the classroom as time, space, and numbers of adults and pupils were omitted from this analysis." [p. 124])

Included in the reliability studies were the 33 factors dervied from the factor analysis of the FLACCS, the TPOR, the RCS and the Cognitive Taxonomy. However, for the GRCD, the reliability study was performed on the original 17 items in the Global Rating instrument.

D. Analysis

The analysis performed to assess stability of observation over time was simply to correlate factor scores of the 17 observed teachers at different pairs of observation times, i.e., time points (1,2), (1,3), and (2,3).

The analysis of the reliability of the 33 factors from the FLACCS, the TPOR, the RCS and the Cog. Taxonomy instruments was not described except for mention that video tapes of classroom activities were used and evidently were rated by different observers, and then analysis of variance was used to obtain some sort of interaction reliability. (pp. 125-126).

The analytical method for the Global Rating items is described as follows:

"Since the classroom description data would not be expected to suffer from problems of reliability (measures such as physical size of the classrooms and numbers of pupils and adults), reliabilities have not been estimated for the GRCD factor scores. The ratings, however, whose reliability is of interest, were recorded separately for each classroom by each member of the observation team. (Recall that each of the 289 classrooms in the Classroom Observation Study was rated by two observers on the GRCD instruments.) Two reliability estimates are reported for each rating; one is the correlation between the two raters, which would be relevant to the use of the ratings by a single observer, the other is a Spearman-Brown estimate of the reliability of the pooled ratings used in further analysis." (p. 106)

E. Comments

Having used the same variables and sample as the "Study of Teacher Behavior at Three Points in Time," the stability analysis is subject to the same caveats as applied to that study. Moreover, having selected two groups on the basis of their extreme initial scores (high and low control by teachers), pooling them together to calculate stability correlations will artificially inflate the stability coefficients for any observations related to the items on which the two groups were selected.

Also, since the stability analysis included only seventeen teachers, statistical significance, although not reported, is not attained by the vast majority of reported results, at either of the commonly report levels of significance (i.e., .01, or .05).

Since the manner of calculation of reliability coefficients for the factor scores is not reported, it is impossible to assess the adequacy of this analysis. Also the author unfortunately deleted certain results for reasons described as follows:

"Large amounts of video tape were screened seeking tapes in which pupil behavior was more prominent, and the tapes used were the most representative that were found, but they still displayed a relatively 'teacher front and center' style of teaching, which meant that several of the measures showed the teachers to be very much alike. Since analysis of variance reliability basically compares variance between observers with variance between teachers, several of these measures showed negative reliabilities, which meant that variance between observers was greater than between teachers. Since these figures depart from the usual meaning of reliability, they are not reported." (p. 126)

The rationale for deleting classroom description variables from the reliability analysis is quite odd. If as suggested, the variables assessed by this instrument "would not be expected to suffer from problems of reliability," the reliabilities might still usefully have been calculated as an indication of observer reliability.

Also the report provides no rationale for why reliabilities are calculated on the basis of item scores in this case, but on the factor scores for the four classroom observation instruments.

F. Results

The reported stability coefficients (Table 104, p. 122) are highly variable, ranging from negative to .87. However, given the small size of the sample (17) and the problems in the gathering of the data for this study (see comments regarding the Study of Teacher Behavior at Three Points in Time), very little, if anything at all, can be concluded from the results of the stability study.

Reliabilities reported for the factor scores (presumably some sort of inter-rater reliabilities) for the 33 factors derived from the four classroom observation instruments range from .48 to .99 with 12 out of 29 equal to .90 or greater (Table 105, p. 127). However, it is impossible to conclude much from these figures because it is not explained how they were obtained.

The correlation between observers ratings or the items in the Global Ratings ranged from .42 to .82 while the Spearman-Brown adjusted pooled reliability coefficients ranged from .59 to .90 (Table 106, p. 128).

From these analyses of the stability and reliability of classroom observations and ratings the Soar report concludes:

"The pattern of relationships across all the instruments and the three periods of time are not generally high, indicating considerable variability of teacher behavior across occasions, even though the analyses of variance did not show large numbers of significant differences. It seems likely that teacher-pupil data would be more variable from activity to activity within a given day than from day to day when the entire day is pooled, as was the case here. The generally higher video and audiotape reliabilities reported in Table 105, p. 127, in comparison to [the stability data], appear to support the suggestion that variability of teacher behavior is a greater source of variance than observer agreement." (p. 124)

"The reliabilities for the FLACCS and the TPOR (Table 105) are generally high; probably some are spuriously high for the reasons cited above [regarding the assessment of observer reliability from video tapes]. (p. 126)

"There was other evidence to indicate that the FLACCS reliabilities which were omitted were at least significantly reliable. All four of them discriminated significantly between experimental programs (Table 102), with probabilities beyond the one percent level. Presumably, this would occur only occasionally, and as a chance occurrence, if the measure did not have significant reliability...

"The TPOR factors generally had satisfactory to good reliability...

"The RCS factors had reliabilities that ranged from questionable to good...

"The reliabilities of the COGTAX were generally surprisingly high...

"The data do illustrate the inadequacy of presenting reliability in terms of observer agreement for a total instrument, which is common. The RCS data, for example, indicate that reliabilities of separate measures ranged from indicating doubtful value to ones that were quite high. To cite one overall value for the instrument as a whole does not seem to be useful." (p. 129)

II. OTHER FOLLOW THROUGH EVALUATION STUDIES FUNDED BY THE OFFICE OF EDUCATION AND THE DEPARTMENT OF HEALTH, EDUCATION AND WELFARE

Included in this section are three types of Follow Through evaluation studies not included in section I. First, are four major Follow Through evaluation reports not summarized in detail in section I; the cost analysis of Follow Through projects by Donald Rogers, et al.; the 1971-72 classroom observation report by Jane Stallings and others; and the Abt III and Abt IV Follow Through evaluation reports by Linda Stebbins, et al. The second type of document summarized in this section includes final reports narrower in scope than the major evaluation studies. The third category includes a variety of interim reports and appendices which never reached final report status, but which nevertheless contain valuable information on Follow Through which is not available elsewhere. Summaries of reports in this section are descriptive rather than critical. Where available, ERIC numbers are provided in order to facilitate access to these materials. Summary descriptions are organized alphabetically by author.

Abt Associates, Inc. *Interim Report*. *Volume I: An Analysis of Selected Data, 1969-1972*. *Volume II: Appendices*. Sponsored by BESE, Division of Compensatory Education, OE, DHEW. Cambridge, MA: Abt Associates, January 31, 1973. 197 pp.

This interim report describes Abt's preliminary analyses of cohort III kindergarten data, and of data for cohort I EF stream children in grades 1 and 3. It was superseded by Marvin G. Cline and others, *Final Report. Education as Experimentation: Evaluation of the Follow Through Planned Variation Model. Volume IA: Early Effects of Follow Through* and *Volume IIB: Monographs*.

Acland, Henry and Marx, Thomas Jakob. *A Study of Teacher Impact on Student Learning*. Cambridge, MA: Huron Institute, 1974. 55 pp.

This report describes analyses of data from approximately twenty classes which had the same teachers over two (or three) years to determine whether some teachers are more effective than others in terms of their students' performance on commercial standardized tests. Issues addressed are: consistency of teacher effects both across tests within years and across years within tests, teacher effects versus sponsor effects, the impact of teacher goals on pupil learning, the problem of test intervals, especially effective teachers, and teachers' effects on the spread of test scores.

Bussis, Anne M. and Chittenden, Edward A. Analysis of An Approach to Open Education: Interim Report. Sponsored by OE/DHEW. Princeton, NJ: Educational Testing Service, 1970. 87 pp. ED 050 125.

This study was guided by two fundamental concerns: the problems of development assessment procedures which are better suited to the more humanistic but less tangible goals of education in general, and the need for a clearer conceptualization of the objectives of open education programs. Two major components of this report are: (1) conceptual analysis of an "open approach to pre-school and primary education; and (2) discussions of implications of this approach for questions of research and evaluation. Particular attention is given to identifying basic assumptions about children's learning, educational change, and the teacher's role. (ERIC summary.)

Crawford, K.D. and Middleton, T. "Comparisons of Responses to Selected Items Contained in Fall 1971 and Spring 1975 Parent Interviews." Menlo Park, CA: Stanford Research Institute, June 1976. 39 pp.

This report examines the responses of 3,455 FT and NFT parents to twelve questions (six main questions, one of which is composed of seven sub-items) who completed parent questionnaires as their children entered Follow Through Fall 1971 and as their children exited Follow Through Spring 1975. Comparison of parent responses to each question at both time points is made. Responses are reported in tables which cross tabulate sponsors with FT/NFT class membership. The topical areas probed by the questions are: parents' sense of self-efficacy, parental satisfaction and expectation concerning their children's education, and school and family life intervention.

David, Jane Lisa. Summer Study: A Two Part Investigation of the Impact of Exposure to Schooling on Achievement Growth. Cambridge, MA: Huron Institute, 1974. 184 pp. ED 124 669. (Revised as a doctoral dissertation, Harvard Graduate School of Education, 1974.)

This study presents an examination of two hypotheses using data gathered from approximately 1000 Philadelphia school children: (1) the difference in achievement growth between poor and non-poor children is greater during the summer months than during the school year and (2) participation in a summer program increases achievement growth and that this increase is maintained throughout the following school year.

Fox, Robert S. An Examination of Support Systems in Selected Follow Through Schools. Washington, D.C.: National Training Laboratories Institute for Applied Behavioral Research, March 1971. 313 pp. ED 068 417.

A study of the effectiveness of various support systems of program models in Follow Through schools examines those characteristics of a support system that strengthen or weaken the program's potential for success. Exploration of the nature of support system variables in a sample of schools is accomplished through group interview techniques for data-gathering. Researchers create a procedure for diagnosing some of the problems with regard to the manner in which these variables are managed and identify or create a set of training activities and designs which have potential value in strengthening support systems in Follow Through programs. These training ideas are organized into a resource manual for further field testing by the participating schools. (ERIC summary.)

Grannis, Joseph C. The Columbia Classroom Environment Project...Second Progress Report. New York, NY: Columbia University Institute for Pedagogical Studies, February 1971. 172 pp. ED 123 288.

This document presents a discussion of processes to be used in recording classroom behaviors in such a way as to be useful for Follow Through evaluation. (First progress report is no longer available.)

Grannis, Joseph C. The Columbia Classroom Environments Project...Third Progress Report. New York, NY: Columbia University, Institute for Pedagogical Studies, May 1971. 146 pp. ED 123 289.

This is a concept paper to serve as the basis for subsequent phases of the CCEP project. The paper focuses on teachers' roles in different classroom settings, teachers' and pupils' nonverbal behavior, procedures for mapping autonomous and integrated pupil behavior, evaluations design, the development of protocols, and a description of a social context study previously conducted.

Grannis, Joseph C. *The Columbia Classroom Environments Project...Fourth Progress Report*. New York, NY: Columbia University Institute for Pedagogical Studies, December 1971. 244 pp. ED 123 290.

This paper discusses a number of questions concerning dimensions of learning and development as well as instruments being prepared for the analysis of behavior in learning environments.

Grannis, Joseph C. *The Columbia Classroom Environment Project...Fifth Progress Report*. New York, NY: Columbia University Institute for Pedagogical Studies, May 1972. 567 pp. ED 123 291.

Using CCEP's August 1970 proposal, the report describes progress in six areas: an analysis of Child Behavior Stream Observations; *BORIS* - a child-based observation scheme to obtain information about children's interactions with teachers, peers, and materials; involvement ratings; and analysis of teacher behavior data; an analysis of social context data; and research procedures in the pupil role interview study.

Grannis, Joseph C. *The Columbia Classroom Environments Project...Final Progress Report*. New York, NY: Columbia University Institute for Pedagogical Studies, February 1973. 170 pp. ED 123 292.

This is a three-part report dealing with (1) the construction and use of a pupil role interview schedule to determine the influence of different Follow Through programs on student role development, (2) the content analysis of a Pupil Role Interview of 480 children, and (3) a theoretical analysis of the control of respective classroom, environmental factors for achieving different educational objectives.

Haney, Walt. *Analysis of Interim Follow Through Evaluation Reports*. Sponsored by OE/DHEW. Cambridge, MA: Huron Institute, September 1976. 177 pp.

This report is a draft version of the 1977 report entitled *A Technical History of the Follow Through National Evaluation*. It differs from the latter study mainly in that it treats the national Follow Through evaluation only through 1975.

Hantman, Stephen and Acland, Henry. Final Report on the Follow Through City Data. Cambridge, MA: Huron Institute, 1974. 54 pp. ED 094 884.

This report critically reviews Follow Through sponsor reports for 1971-72 and points up the great variation in sponsor goals--a variation that produces difficulties for standardized evaluation. A second section examines test findings from "structured" sponsors (Oregon, Kansas and Pittsburgh) and discusses the possibility that differences in child performance scores may be more the product of the type of test administered than of differences in respective models. A third section compares Follow Through schools with other "disadvantaged" schools by noting differences between 1971-72 Follow Through student performance scores and scores of children in other New York and Baltimore schools.

Hardy, Robert and Huebner, Robert. Evaluation of Follow Through Programs by Follow Through Sponsors 1970-71. A Summary, Analysis, and Review. Sponsored by the Bureau of Elementary and Secondary Education, OE/DHEW. December, 1971. 129 pp. ED 107 369.

This report contains a summary and analysis of the evaluation component of Project Follow Through Programs by the 1970-71 Follow Through sponsors. Part 1 (the major part of the document) is a sponsor-by-sponsor summary analysis of the 19 Follow Through Model Programs. Each summary analysis is divided into five parts: nature of the program, process measures (designed to assess processes used to accomplish pupil growth), product measures (designed to assess any aspect of pupil growth), conclusions of the sponsor, and reviewer's comments. Part 2 consists of a coded listing of the process and product variables of each of the 19 model programs and summaries of some of the program findings. Part 3 is a presentation of measures on a matrix, with the sponsors along one dimension of the matrix and the common classification system along the other dimension. (ERIC summary.)

Hargroves, Jeannette. Final Report on the Pupil Assignment and Promotion Study. Cambridge, MA: Huron Institute, 1974. 73 pp. ED 094 885.

The study examines Follow Through effects on promotion using data from 22 sites in Follow Through cohorts I and II and information from end-of-school-year promotion lists for four Follow Through sites.

Jackson, Gregory. "Review of Follow Through Sponsor and Local Data." Preliminary Report (15 May 1974), 20 pp. Interim Report (30 July 1974), 54 pp. Final Report (31 August 1974), 33 pp. Sponsored by OE/DHEW. Cambridge, MA: Huron Institute.

These three reports describe Huron's work in reviewing "...sponsors' and local evaluation reports...to assist OPBE/OE in the development of criteria for identifying successful Follow Through models." The preliminary report describes the possibility for "triangular" evaluation of Follow Through projects; as viewed through national sponsor and local reports. The interim report analyzes the data found in 106 reports (by 85 sites and 21 sponsors) for the 1972-73 school year. These reports were found to be widely varied in terms of test instruments; sampling procedures, test score transformations; variability estimates; and correspondence with sponsors' goals. As a result, the report indicates these reports would be of limited usefulness for judging program effectiveness in different sites. The final report discusses general issues in program evaluation and dissemination and suggests alternative criteria for identifying successful Follow Through models in light of data available. It is concluded that determination of criteria of success depends on clarification of program goals; and that "based on the information available so far, program validation will not be a rigorous technical procedure. ...[since] The data which sponsors and sites submit in the absence of guidance or requests is simply not suitable to evaluation the overall program."

Jesse, Ronald N. "The Follow Through Program, 1971-72 Regular and 1972 Summer School Terms. Consolidated Program Information Report." Washington, D.C.: National Center for Educational Statistics/ DHEW, 1975. 5 pp. ED 124 610.

This is a statistical description of Follow Through programs operating during the period Fall 1971 through Summer 1972 based on a survey of 150 agencies who sponsored or managed the various programs.

Judd, David E. and others. "An Introduction to the Follow Through Implementation Study." Sponsored by OE/DHEW. Portland, OR: Nero and Associates, January 15, 1974. 19 pp.

This booklet describes plans for NERO's study of the implementation strategies of Follow Through sponsors.

Judd, David E. and others. Project Follow Through: A Description of Sponsor Implementation Processes. Portland, OR: Nero and Associates, April 1975. 220 pp. ED 128 926.

This report documents Follow Through sponsors' and school districts' experiences in implementing Follow Through programs during the first seven years of Project Follow Through's operations. The report focuses on teacher-child interactions characteristic of respective program approaches, teacher-trainer and trainer-staff interactions, formative and summative evaluation, factors inhibiting or facilitating adequate program administration.

Judd, David E. and Wood, Sheryl E. Follow Through Materials Review. A Description of Selected Materials Developed and Used in Each of Twenty Follow Through Models. Sponsored by OE/DHEW. Portland, OR: Nero and Associates, June 1973. ca. 350 pp.

This report is a description of twenty of the twenty-two Follow Through models. The descriptions of each model focus on: identity of the sponsor and the extent of its program, the philosophy underlying the sponsor's model, classroom instruction, staff development, parent education/involvement, and evaluation system.

[Kennedy, Mary.] Preliminary Review of Sponsor and LEA Evaluation Data in Follow Through. A Working Paper. Washington, D.C.: Office of Planning, Budgeting and Evaluation, Elementary and Secondary Division, OE/DHEW, 1975. 71 pp.

This draft document reviews problems of comparing and synthesizing Follow Through sponsor and LEA evaluation data. Differences in types of comparison groups employed; differences in grades evaluated; and differences in tests and test metrics used are all discussed. The report notes that in an effort to surmount these sorts of problems in sponsor and LEA evaluation activities, a guide for analyzing and presenting evaluation data and an outline for preparing Follow Through evaluation reports was sent to all Follow Through sponsors and LEA's.

Krulee, Gilbert K. and others. An Organizational Analysis of Project Follow Through. Final Report. Evanston, IL: Northwestern University, 1973. 413 pp. ED 093 446.

The purpose of this study is to provide an assessment of the Follow Through organization and its major components in order to identify problems in functioning and to suggest possible organizational changes that would lead to improved functioning. Each of the major components of Follow Through - the Follow Through Office, the sponsors, local projects, and the evaluation effort - is described and the organization of these components into a total program is outlined. The historical development of the program as well as its present structure and functioning is reviewed. Conclusions and implications are drawn from the Follow Through experience that would appear to be relevant to the administration of similar programs of intervention and experimental innovation. (ERIC summary.)

Krulee, Gilbert K. and others. Some Guidelines for Organizing a Local Follow Through Project. Sponsored by the Bureau of Elementary and Secondary Education, OE/DHEW. Evanston, IL: Northwestern University, 1973. 87 pp. ED 089 841.

This report summarizes the experiences of existing Follow Through projects in an effort to provide some guidance for communities attempting to establish new Follow Through programs. Problems commonly encountered in setting up project objectives (choosing a sponsor, recruiting and training staff, establishing liaison between project, school, sponsor, and community and monitoring project effectiveness), are described, and lessons to be learned from both successful and unsuccessful experiences are reviewed. (ERIC summary.)

Kurfeerst, Marvin and Stephens, Thomas M. The National Evaluation of Project Follow Through: 1967-68. Pittsburgh, PA: Pittsburgh University Office of Research and Field Services, May 1969. 329 pp. ED 078 923.

This document reports on the first national Follow Through evaluation effort: an evaluation of 29 Follow Through project sites based on observations and data from the beginning of the 1967-68 school year.

Madaus, George F. and Marx, Thomas J. Report on the Follow Through Later Grades Study. Sponsored by OE/DHEW. Cambridge, MA: Huron Institute, 1973. 64 pp.

This study reports reasons for conducting a study of Follow Through children in upper elementary grades after they leave the Follow Through program, and discusses issues that need to be faced in planning a Follow Through later grades study.

Middleton, Teresa and Wright, Anne R. Report on Parent Interview Pretest Findings. Menlo Park, CA: Stanford Research Institute, August 1976. 99 pp.

"This report deals with the design, conduct, and results of the pretest of the parent interview instrument" (p. 5). This instrument was being developed for use in a parent study which would examine the responses of Cohort III parents as their children exited from third grade and at time points during their children's attendance in fourth and fifth grades. Some of the data collected during this pretest are compared with data collected in spring 1975.

Miller, Louise B. and Dyer, Jean L. Two Kinds of Kindergarten After Four Types of Head Start. Sponsored by the National Institute of Child Health and Human Development. Louisville, KY: Louisville University, 1971. 52 pp. ED 050 824.

This document reports the evaluation results of a three-year comparison study of children with four types of pre-kindergarten experience who had subsequent Follow Through experience. Implementation issues are discussed along with interaction effects and sex differences.

National Training Laboratory, Institute for Applied Behavioral Science. Revised Final Report. Facilitating Effective Involvement of Parents, Staff, Consultants, Sponsors and Community Leaders in Project Follow Thro Program Development at the Local Level. July 1, 1970-September 30, 1972. Arlington, VA: NTL Institute for Applied Behavioral Science, March 1976. 100 pp.

The report describes the goals, administrative network, and methods used to promote the more effective involvement of Follow Through Policy Advisory Committees, sponsors, parents and community groups, school administrators, and school board members in Follow Through activities. The report describes NTL's efforts in ten school districts during 1970-71 and 1971-72.

Rath, Susan W. and others. Follow Through: A Story of Educational Change. Sponsored by OE/DHEW. Portland, OR: Nero and Associates, 1976. 135 pp. ED 129 418.

This survey of Follow Through programs over the past nine years examines program implementation from the perspective of local schools (complementing an earlier study on implementation of education models from the perspective of model sponsors). Ten sites around the country were visited and over 300 Follow Through participants interviewed, including parents, children, teachers, paraprofessionals, principals, Follow Through staff, administrators and school board members. Dealing with Follow Through program design and evaluation only as related to implementation, the study focuses on how Follow Through is translated into day-to-day operations. A study team searched for "lessons learned" that could help others toward successful implementation of programs. Key chapters discuss the implementation process as it relates to: changing the classroom, parent involvement, support services, career development, and changes in the schools. (ERIC summary.)

Resnick, Lauren B. and others. Follow Through: Program Approaches, School Year 1970-71. Washington, D.C.: Office of Economic Opportunity, 1970. 51 pp. ED 047 787.

A general description of the approaches used by each of the twenty models operating in 160 public schools during the 1970-71 school year.

Rogers, Donald D., et al. A Cost Analysis of Follow Through Projects. Bethesda, MD: RMC Research Corporation. April 1977. 96 pp. text and ca. 150 pp. appendices.*

RMC's cost analysis is based on information obtained from (1) project directors, principals, and teachers at 51 FT sites which included 147 schools and 930 classrooms and (2) principals and teachers at 44 NFT sites which included 262 classrooms. Information was collected during the 1975-76 school year.

A "resource approach" to cost analysis was employed. Utilizing information obtained from the questionnaires 114 types of resources used in classrooms were identified. Estimates of the quantity of each type of resource used in a classroom were made and each multiplied by a standard unit cost to produce a cost per classroom estimate. Cost per class is the unit of analysis employed in most of the commentary, but some observations dealt with school and child unit costs as well. Analyses produced findings in three general areas.

---

*This report became Volume IV in OE's final report The Follow Through Planned Variation Experiment.

Follow Through/Non-Follow Through Comparisons. FT classroom costs were compared with comparison programs located within the same school district. Comparisons were made for 28 aggregated resources categories. Overall differences and significant differences within each of these categories were noted. Typically speaking FT classrooms were found to cost $19,000 more than comparison classrooms per year. Resources accounting for the difference between FT and NFT class costs vary from sponsor to sponsor but in general 46% of the average cost difference is due to school staff, primarily instructional aides and staff/curriculum developers.

Implementation Costs. The cost of implementing a FT program in a classroom varied considerably both from sponsor to sponsor and from site to site. Moreover, costs of implementing NFT comparison classrooms also varied from site to site.

The probable total cost of implementing a sponsor's program ranged from $1374 to $2689 per pupil. The average total cost in a NFT program was reported as $1282 per pupil. Per pupil, the total cost differences between FT and NFT classrooms were thus found to range from $92 to $1407. This difference was primarily attributed to differences in economic conditions at sites where programs were implemented

Model to Model Comparisons. Sponsors were compared on the basis of averaged within-site, FT-NFT incremental cost differences. Comparisons were drawn for 28 resources categories. Significant inter-sponsor differences were identified for one-half the resource categories. Categories contributing the largest percentages to incremental cost differences were: sponsor; staff/curriculum developer; parent coordinator/home visitor; nutrition; volunteers; and equipment, materials and supplies. These findings were strictly qualified as preliminary and tentative due to the probability of bias introduced by the observed dependency of cost differences on site conditions, a factor not adequately controlled for in the sample design.

Soar, Robert S. Follow Through Classroom Process Measurement. Gainesville, FL: Florida University Institute for Development of Human Resources, 1971. 177 pp. ED 113 288.

Using data from students from seven Follow Through models and one comparison site this evaluation report describes the differences among programs as observed in the classroom and attempts to relate these differences to pupil growth.

This report focuses on the first two years of preliminary work by Soar and his associates at the University of Florida, The final report on their work is summarized in section A of this bibliography.

Stallings, Jane A., et al. Follow Through Program Classroom Observation Evaluation 1971-72. Menlo Park, CA: Stanford Research Institute, August 1973. 547 pp. ED 085 100.

This report describes results of inquiries into: (1) program implementation--the degree to which classroom processes reflect a given sponsor's model and (2) the relationships between classroom processes and child test outcomes.

One project site was selected for each of twelve sponsors. A total of 146 FT and 74 NFT classrooms--including kindergarten, first, second, and third grades--were observed during the spring of 1972.

Data from a Five-Minute Observation instrument were examined to determine the extent of activity in ten areas--e.g., math, reading, domestic arts, blocks and trucks. Within grade level means and standard deviations for each sponsor were calculated. Similarly, Classroom Check List data were used to calculate means and standard deviations for the relative frequency of teacher working with one child, two children, a small, or a large group on an academic activity. Within grade level means and standard deviations were obtained for children working without the teacher by themselves, with another child, with a small or with a large group. Similar calculations were made to describe the extent of adult feedback for child behavior, for task-related activity, and overall. And, finally, the physical environment of each sponsor's typical classroom was described on the basis of information on a Classroom Observation Instruments. These measures form the basis for a description of the sponsor's average classroom.

Implementation. Site implementation scores were devised in an effort to determine whether respective sponsors had been successful in implementing their model. From a wide range of classroom descriptors, a set of variables deemed appropriate for the description of what should occur in a given sponsor's classroom was selected for each sponsor. The number of variables in each sponsor's set ranges from fourteen for the University of Florida to 43 for Far West. NFT classroom scores on selected variables were distributed and divided into quartiles. FT classroom scores were averaged to produce a FT site level mean. FT site level means were compared to the NFT score distribution. Each site level mean was assigned a score of 1 if it fell within the range of the lowest quartile of NFT scores--a score of 4 if it fell within the range of the highest quartile of NFT scores. The resulting values were summed across all the variables of the set selected to describe a particular sponsor's model and divided by the maximum that could be obtained (4 times the number of variables used to describe a sponsor's model). The resulting site implementation scores were expressed as percentages and ranged from 77.8% for Southwest Lab to a high of 93.2% for the University of Oregon. The report concludes that differences in site implementation scores may indicate that implementation of behavior modification models (University of Oregon, Kansas and Pittsburgh) are more readily measured by the observation instruments used.

Sixty-five variables were selected from a universe of 200 classroom descriptors and subjected to factor analysis. A nine-factor rotation accounted for 57% of the variance in the matrix. These nine factors--stimulus-response-feedback, small group activities, range of emotion in social behavior, child initiative, formal instruction, divergent questioning, individualized work setting, academic equipment and materials, and children not engaged with adults--provide the basis for inter-sponsor comparisons and for an overall FT/NFT comparison. Intersponsor differences were many and varied. Overall FT/NFT comparisons indicated no or negligible differences on stimulus-response-feedback, child initiative, formal instruction, and child not engaged with adults. FT scores were noticeably higher on small group activities, divergent questioning, individualized work setting, and academic equipment and materials. A discernably higher NFT score was obtained on range of emotion in social behaviors.

Twenty-five classroom descriptors were selected on the basis of their having been used in both the 1971 and 1972 observations. (Comparable variables were constructed in some cases.) Sponsor scores on these 25 variables were ranked for each year and the rankings for each year were rank-order correlated. High correlations were obtained for: games and puzzles; blocks, trucks, dolls, and dressup; wide variety of activities; child questioning adult; and adult praise or acknowledgment of children.

Overall FT/NFT differences on 23 Observation Summary Form variables, 55 Classroom Check List variables, and 120 Five-Minute Observation schedule variables were also tested. Variables for which significant differences were found were classified into five categories; those having to do with: (1) groupings of teachers, aides, and volunteers with children; (2) classroom activities; (3) child independence; (4) equipment and materials used; and (5) interaction and behavior of adults and children in the classroom. In general, FT teachers and aides more often work with individuals or a small group of children during reading and arithmetic learning activities. FT classrooms have a wider range of activities using a wider range of equipment. Although there is greater independent activity in FT classrooms, adults are more likely to initiate interactions with children. While NFT teachers showed a tendency to instruct or lecture students more, FT teachers tended to use concrete objects as part of the instructional process more often. FT children were found to respond, initiate interaction with adults, instruct themselves, and instruct other children more often. The amount of positive feedback given by both FT and NFT adults for non-affective child behavior did not differ significantly. FT children received more praise for good behavior; NFT children received more negative corrective feedback for misbehavior.

In an effort to develop hypotheses for subsequent investigations into the nature of the relationship between classroom procedures, an exploratory study of the relationships between adult behavior and child behavior was undertaken. Attributes considered important in child growth and development, i.e., verbal initiative, questioning, responsiveness, independence, cooperativeness and positive behavior, were selected as salient child behaviors. Six variables taken from the child-focused portion of the Five-Minute Observation schedule were used as measures of these child behaviors. Scores on these variables were correlated with scores obtained on adult activities obtained from the adult-focused portion of the Five-Minute Observation schedule. Sizeable correlations were noted and discussed.

Process and Child Test Outcomes. Students for whom all outcomes measures and entry WRAT scores were available were the subjects of several analyses which attempted to identify the nature of the relationship between classroom processes and child test outcomes. In all 1044 FT/404 NFT entering kindergarten and 1041 FT/421 NFT entering first grade students in grades k through 3 were included.

Children in different grade level/streams were administered different combinations of test items (or different combinations of complete tests) from eight instruments: Wide Range Achievement Test, Metropolitan Achievement Test, Peabody Picture Vocabularly Test, Raven's Coloured Progressive Matrices, Gumpgookies, ETS Version of Locus of Control, Intellectual Achievement Responsiblity Scale, and Coopersmith Self-Esteem Inventory. Seven variables were used in an effort to adjust for entry level differences: age, sex, ethnic origin, months of Head Start experience, months of Follow Through experience, days absent, and entry WRAT scores. The child was the unit of analysis. Since different tests were administered at different grade level/streams, various analytical methodologies were applied to data from each grade level/stream separately.

Analysis of covariance was used to obtain coefficients of determination for the regression of child test scores on all variables and on all variables plus sponsors. Results indicated that covariables accounted for a good proportion of the variance in respective academic test scores but only a small amount in the affective test scores. The increment to the coefficient of determination achieved by adding sponsors as independent variables proved to be significant in nearly all cases. Exceptions to this general observation were found with affective test outcomes and occasionally with Raven's or Coopersmith test scores.

Sponsor differences were likewise calculated using an ANCOVA contrast design. In this analysis, the means of various sponsors' students are contrasted with an overall mean sponsor effect, and consequently with one another, after adjusting for the effects of covariates. NFT means on outcomes are included in the contrasts. Results are reported but not discussed.

Twelve adult-focused and ten child-focused classroom descriptors were selected from items on different sections of the classroom observation instrument. These measures of classroom processes became independent variables for another series of ANCOVA analyses. The same child test scores used as dependent variables, and the same set of covariates, as were used in the contrast analysis of sponsors effects were used in this analysis of classroom process effects. Child test scores were regressed on the set of covariables and then on the covariables and the set of 22 classroom process variables to generate coefficients of determination. Results indicated that (1) covariates account for a considerable amount of variation in child academic test scores, (2) the amount of additional explained variance achieved through the addition of classroom process variables into the regression was almost always significant, and (3) the classroom process variables usually account for approximately 13% of the variance in academic achievement scores.

A correlation of student test scores with scores on nine factors identified by the implementation analyses as describing classroom processes was undertaken. The correlations did not show strong relationships between student test scores and scores on these factors.

Child test scores were also correlated with a set of 26 classroom observation variables using partial correlation analysis. Partial correlations greater than an absolute value of .15 were identified. High positive correlations were discussed under three headings--academic activities, child participation, and adult response to child. Large negative correlations generated by the partial correlation analysis were discussed under four headings: child initiation and instruction, child waiting, wide variety of activities; and unexpected findings.

In another effort to examine the relationship between classroom processes and child test scores, four classroom process variables reflecting the amount of individual attention given to students were correlated with student test scores. The trend in sizeable coefficients obtained was interpreted to suggest that benefits accompanying individual attention follow the pattern of the child's learning processes; the child benefits most from individual attention when that attention is appropriate to his developmental needs.

Sponsors were classified according to the primary goals they promoted: self-actualizing sponsors promoting socioemotional goals (Far West, University of Arizona, Bank Street, and EDC); behavior modification sponsors promoting traditional academic goals (the Universities of Oregon, Kansas, and Pittsburgh); and cognitive growth sponsors emphasizing concept development through experience (Far West, University of Georgia, High/Scope and ILM). ANCOVA-generated adjusted effects were used to determine (1) whether students in models that promoted the development of traditional academic skills would score higher on academic tests than students in models oriented toward socio-emotional goals and conversely (2) whether students in classrooms of the latter type would do better on affective tests than students in classrooms of the former type. The results of the analysis suggested that students in academically oriented programs do better on academic tests than do those in programs emphasizing socio-emotional objectives. However, students in the latter type of program do not consistently do better than those in academic programs on tests measuring affective outcomes. ANCOVA-generated adjusted gain scores were also used to determine whether (1) students in classrooms of sponsors who emphasized concept development would do better on tests of problem solving skills (Peabody and Raven's) than students in the academically oriented classrooms and conversely (2) whether the former would do less well than the latter on academic achievement tests. Results suggested that students in programs emphasizing concept development did not do as well as those in academically oriented programs on academic achievement tests, but did seem to do better on problem solving tests (though the latter contrasts did not generally attain the level of statistical significance).

Stanford Research Institute. "Consolidated Results of the Interim Evaluation of the National Follow Through Program: 1969-1971" Revised Version. Menlo Park, CA: Stanford Research Institute, March 13, 1973. 20 pp.

This report summarizes in bar graphs and very brief summary statements the SRI *Interim Evaluation of the National Follow Through Program 1969-1971: A Technical Report* by John A. Emrick, Philip H. Sorenson and Marian S. Stearns. Stanford Research Institute, February 1973.

Stanford Research Institute. *Evaluation of the National Follow Through Program 1969-1971*. Draft version. Menlo Park, CA: Stanford Research Institute, August 1972. 300 pp.

This is SRI's 1972 report on Follow Through program effects on children, parents and teachers. Effects on children are discussed in terms of a battery of cognitive and non-cognitive measures. Effects on parents are described in terms of parent-child interaction, parent-school interaction, parent expectations and degree of satisfaction concerning child performance, and parent sense of control over school procedures, reforms, and compliance to community desires. Effects on teachers are examined in terms of teachers' sense of need to work with parents and their commitment to the sponsor methods they are using.

Stanford Research Institute. *Evaluation of the National Follow Through Program 1969-1971. Annexes I-VI*. Draft version. Menlo Park, CA: Stanford Research Institute, August 1972. 88 pp.

These six annexes to the August 1972 evaluation report cover the following topics: (1) Item Analysis and Reliability Data for Pupil Outcome Measures, (2) Operational Description of Variables, (3) Regression of Control Variables on Outcome Measures for pupil, parent, and teacher analyses, (4) Results of Project-Level Analyses on Disaggregated Achievement Variables (quantitative, reading, language, and cognitive processes), (5) Tables of Unadjusted Means and Standard Deviation on Pupil Outcome Measures, and (6) Special Regression Analysis (path model of Follow Through and Child Test Dependent Variables).

Stanford Research Institute. Evaluation of the National Follow Through Program 1969-71. Appendix A: Follow Through Program Sponsors. Draft version. Menlo Park, CA: Stanford Research Institute, August 1972. 69 pp. ED 080 494.

This document describes each of the 22 Follow Through sponsor models and the nature of their programs.

Stanford Research Institute. Evaluation of the National Follow Through Program, 1969-1971. Appendix B: Follow Through Classroom Observations. Draft version. Menlo Park, CA: Stanford Research Institute, August 1972. 134 pp.

This technical appendix describes the development of the Classroom Observation Instrument used by SRI to describe (1) the amount of time devoted to specific activities, (2) the organization of classroom learning groups, and (3) the amount and kind of communication in the classroom. Data obtained by its application in NFT classrooms and in selected classrooms of eight FT sponsors during the 1970/71 school year are used to describe FT versus NFT classroom differences and variations among sponsor classrooms. An effort is made to relate classroom processes to outcome measures.

Stanford Research Institute. Evaluation of the National Follow Through Program 1969-1971. Appendix C: Follow Through Pupil Tests, Parent Interviews, and Teacher Questionnaires. Draft version. Menlo Park, CA: Stanford Research Institute, August 1972. 109 pp.

Test batteries used in 1968, 1969/70, and 1970/71 for measuring student performance are described. The development of parent interviews and a teacher questionnaire is described in terms of their purpose, the specification of variables to be measured, the history of the instrument, the selection of samples, field operations, and the derivation of respective scales.

Stanford Research Institute. Evaluation of the National Follow Through Program 1969-1971. Appendix D: Follow Through Project Descriptions. Draft version. Menlo Park, CA: Stanford Research Institute, August 1972. 70 pp.

This appendix describes SRI efforts to provide a comprehensive and systematic description of Follow Through projects (and the diversity of social contexts in which they occurred) for the purposes of program management and planning and to provide data with which to determine overall impacts of Follow Through. Data were collected on the community, the school system, and the project in order to describe the social context of each project. Geographic, demographic, and economic data were also collected.

Stanford Research Institute. Evaluation of the National Follow Through Program, 1969-1971. Appendix E: Follow Through Community Studies. Draft version. Menlo Park, CA: Stanford Research Institute, August 1972. 79 pp.

This appendix describes SRI's efforts to examine the reciprocal influence between the Follow Through program as implemented and existing institutional structures in project communities in order to identify and understand emerging patterns of institutional change occurring within the schools and in the schools' relationships with their communities. A theoretical parent participation model is specified. Path analysis is used to explore the relationships with the seven variable models. Regression results are presented and discussed.

Stanford Research Institute. Evaluation of the National Follow Through Program 1969-1971. Appendix F. Administrative History of the Follow Through Evaluation, 1968-1972. Draft version. Menlo Park, CA: Stanford Research Institute, August 1972. 56 pp.

This appendix describes "the evolution of decisions that affected evaluation design, analyses, and reporting" in the period July 1968 to June 1972. "The emphasis...is upon occasions in which decisions were made or principles were established or reaffirmed" (p. F-1).

Stanford Research Institute. Implementation of Planned Variation in Head Start: Preliminary Evaluations of Planned Variation in Head Start According to Follow Through Approaches (1969-1970). Interim Report: First Year of Evaluation, Part II. Menlo Park, CA: Stanford Research Institute, May 1971. 446 pp. ED 052 844.

This study presents preliminary evaluation results based on analysis of first year data. The report describes the philosophical differences among eight HSPV models and a battery of fourteen measures to record cognitive effects. It also discusses the organization of classrooms, parent-child interactions, child performance data, and diffusion effects.

Stanford Research Institute. Longitudinal Evaluation of Selected Features of the National Follow Through Program. Draft version. Menlo Park, CA: Stanford Research Institute, March 1971. 231 pp. and 135 pp. ED 057 266.

An evaluation is made of the efficacy of various approaches to early childhood education when tested in the natural laboratories of operating schools, and the Follow Through program's impact on achieving social changes and ameliorating effects of poverty through education. Judged primarily from data collected during the 1969-70 school year, Follow Through is achieving its intended objectives. Some of the major findings of the program evaluation are: (1) Effects of Follow Through participation on children, parents, and school staff are not identical from one sponsored approach to another; (2) Children at all grade levels in Follow Through showed greater gains in school achievement than did their non-Follow Through counterparts; (3) Follow Through children showed positive changes in their attitudes toward school, their teachers, and their classmates; (4) Parents of Follow Through children showed consistently higher levels of awareness of their children's school program; were more likely to visit school; tended to be somewhat more satisfied with their child's school; and regarded the school's program as being more efficacious; and (5) Teachers and other school staff generally viewed the program as very helpful to the children and as influencing their teaching practices and attitudes. There are 8 chapters, 8 numbered appendices, 12 illustrations, and 31 tables in this report. (ERIC summary.)

Stanford Research Institute. Longitudinal Evaluation of Selected Features of the National Follow Through Program. Appendix A: Statistical Data. Draft version. Menlo Park, CA: Stanford Research Institute, March 1971. ED 057 267.

This appendix contains statistical data supporting the 1969-70 progress report, "Longitudinal Evaluation of Selected Features of the National Follow Through Program." It is organized as follows: Part 1: Tables A-1 through A-4 show distributions of Fall 1969-70 achievement scores (expressed as percent correct) by grade for Follow Through and Non-Follow Through pupils; Part 2: Tables A-5 through A-70 show Spring and difference scores for the achievement, attitudes toward school, and interpersonal feelings measures by grades and sponsor group for Follow Through and Non-Follow Through pupils; Part 3: Tables A-71 through A-109 show achievement, attitudes toward school, and interpersonal feelings scores by grade and according to preschool experience, poverty classification, and Follow Through services received; Part 4: Tables A-110 through A-118 show scores for nine non-cognitive variables by grade and sponsor group for Follow Through and Non-Follow Through pupils; Part 5: Tables A-119 through A-127 show direction of differences between Follow Through and Non-Follow Through kindergarten and entering first grade pupils on measures of achievement, attitudes toward school, and interpersonal feelings in relation to various measures of behavior and attitudes in which Follow Through and Non-Follow Through parents differed. These tables are also organized by sponsor group. (ERIC summary.)

Stanford Research Institute. *Longitudinal Evaluation of Selected Features of the National Follow Through Program. Appendix D: Issues in Noncognitive Measurement*. Menlo Park, CA: Stanford Research Institute, December 1970. 127 pp.

This appendix is an appendix to the December 1970 version of the *Longitudinal Evaluation of Selected Features of the National Follow Program*. That draft was superseded by a March 1971 draft. Part of the revision consisted in dropping this appendix from the report.

This study reports the results of a 1970 pilot study into the non-cognitive effects of Follow Through, i.e., the social, emotional and personal areas of pupil growth and development. Instruments, the rationale for their selection, the procedures used in the pilot study, results, and recommendations concerning future non-cognitive effects evaluation are presented and discussed.

Stanford Research Institute. *Longitudinal Evaluation of Selected Features of the National Follow Through Program. Appendix H: Program Descriptions*. 11 Volumes. Draft version. Menlo Park, CA: Stanford Research Institute, September 1970.

The Longitudinal Evaluation of Selected Features of the National Follow Through Program appears to be the second, large-scale evaluation report submitted by SRI. This report went through several revisions. We can identify at least three bearing the same title: September 1970, December 1970, and March 1971. *Appendix H: Program Descriptions* was an appendix to the September 1970 draft. Revisions in the September draft led to its being dropped from subsequent drafts. It is included here because it is the only report of information derived from systematic interviews of Follow Through coordinators and Policy Advisory Committee Chairmen we have been able to identify.

The Appendix contains eleven volumes with the following titles:
Volume I - Child Roster Data
Volume II - Personnel Roster Data
Volume III - Policy Advisory Committee Roster Data
Volume IV - Follow Through Coordinator Interview, Part I
Volume V - Follow Through Coordinator Interview, Part II
Volume VI - Follow Through Coordinator Interview, Part III
Volume VII - Follow Through Coordinator Interview, Part IV
Volume VIII - Policy Advisory Committee Chairman Interview, Part I
Volume IX - Policy Advisory Committee Chairman Interview, Part II
Volume X - Policy Advisory Committee Chairman Interview, Part III
Volume XI - Index

These volumes contain a wide range of information regarding *inter alia*: how Follow Through schools were selected; how Follow Through children were selected; coordinator and PAC chairman impressions concerning satisfaction with sponsors' training programs; teacher and parent participation at PAC meetings; parent contributions to Follow Through programs; major Follow Through student, parent, and teacher successes; difficulties with evaluations; selection, formation, and attendance at Policy Advisory Councils; the nature of PAC activities.

Stanford Research Institute. *Materials for Report to Congress on the National Evaluation of Follow Through: 1969-1971. Interim Report.* Menlo Park, CA: Stanford Research Institute, January 1972. 80 pp.

"This report contains selected preliminary findings from evidence gathered during the period Fall 1969 through Spring 1971" (p. 1). Preliminary findings are reported on two groups of children: a Kindergarten Cohort and a First Grade Cohort. All Children included in the analysis are those categorized as "poor" by OEO standards. Primary focus is given to pupil achievement as measured by testing.

Stanford Research Institute. *A Preliminary Guide to Follow Through. A Progress Report on the National Evaluation: 1969-1971.* Menlo Park, CA: Stanford Research Institute, 1972. 100 pp.

This report describes the Follow Through program in terms of program sponsors, project characteristics, and an overview of the evaluation plan. Preliminary evaluation findings are discussed with reference to classroom processes, characteristics of Follow Through teachers, parent reported behavior and beliefs, and pupil performance.

Stanford Research Institute. Richmond Virginia (1005) Feedback Data for Spring 1975: Roster Data, Test Data, Teacher and Aide Survey Data, Parent Survey Data. Menlo Park, CA: Stanford Research Institute, undated, ca. 500 pp.

This document is one of hundreds of such feedback reports sent by SRI to personnel of Follow Through projects involved in the national evaluation on all data collected at those projects in each of the years 1972, 1973, 1974 and 1975.

Stebbins, Linda B., et al. *Education as Experimentation: A Planned Variation Model*. *Volume IIIA. Findings: Cohort II. Interim Findings: Cohort III* and *Volume IIIB. Appendices*. Cambridge, MA: Abt Associates, Inc., July 1976. 407 pp. and 428 pp.

This is the third in the series of Abt Follow Through evaluation reports. It reports analyses of data for cohort II entering kindergarten children who had completed third grade, and cohort III entering kindergarten children who had completed second grade. The main focus of the report is cohort II children and their test data after a full four years of Follow Through. The report has three main parts.

The first part (chapter 1) of the Abt III report is introduction. It recounts the background to the Follow Through planned variation experiment and describes the nature of the FT planned variations. This is followed by a brief summary of the Follow Through evaluation and three main problems the evaluation has faced: 1) lack of specificity of treatment, 2) limitations of the measurement battery, and 3) weaknesses in the design of the evaluation.

Part two has three chapters describing the focus of the Abt III report in terms of: chapter 2) educational strategies of the sponsors included in this evaluation report and the test outcome measures employed; chapter 3) the characteristics of the sample analyzed; and chapter 4) the strategies used in analyzing the cohort II third grade and cohort III second grade data. The report encompasses data from children served by fourteen Follow Through sponsors (including self-sponsored as a single type of sponsor). For descriptive purposes these sponsors are cross-classified according to the degree of structure of their models (high, medium or low), their theoretical emphasis (behaviorist, cognitive or psychodynamic), emphasis on the home environment (high, medium or low) and utilization of parents as teaching agents (high, medium or low). Analyses of parent interview data and teacher questionnaire data are reported to provide indications of acceptance of model tenets. Follow Through outcome measures (MAT, Ravens, Coopersmith and IARS) are explained in terms of sponsor goal areas (basic skills, cognitive conceptual skills, and affective measures). The analytical sample used in this study is then described in terms of how cases were selected and according to several background measures.

Chapter 4 opens with a discussion of six analytical problems troubling the Follow Through evaluation (attrition, missing data, level of analysis, unit of analysis, sample sizes, noncomparability of treatment and comparison groups). Approximately 50% of the FT and NFT children tested at entry into Follow Through were not present for testing at the end of third grade. Missing data problems were handled by including missing data codes in analyses performed. Based on previous evaluation findings, the level of analysis chosen for this study was the site and the unit of analysis the child. The presence of non-comparable groups of children was the prime reason for selection of a three-pronged analysis strategy using 1) local comparisons, 2) best-match comparisons and 3) pooled NFT group comparisons.

The three comparisons each used a type of ANCOVA, although slightly different sets of covariates and different but overlapping samples of children were used in each. Therefore the assumptions underlying the ANCOVA model are considered and special attention is given to problems of 1) differential preschool attendance, 2) adequacy of adjustment procedures to compensate for initial mismatch and 3) problems of knowing which adjustment procedures to trust.

Part three consists of fourteen chapters (one for each sponsor) of results of the analyses. Each sponsor chapter is comprised of: 1) overview; 2) description of sponsor's model; 3) description of learning environment (that is, results of analyses of parent and teacher data for the sponsor); 4) characteristics of the sample analyzed for the sponsor; 5) estimated effects on children, by site; 6) preliminary comments on Cohort III effects to date (where available); and 7) summary.

The Abt III report contains no overall summary or conclusions. From a substantive point of view, the Abt III report was almost completely superseded by the Abt IV report because the latter study included results from the analyses reported in the Abt III study. Nevertheless, the Abt III report is of considerable interest in terms of how the Follow Through evaluation analyses developed and changed over time. Volume B of the Abt III report is comprised entirely of data tables and figures presenting effects estimates by site and by outcome measure.

Stebbins, Linda B., et al. Education as Experimentation: A Planned Variation Model. Cambridge, MA: Abt Associates, April 1977. Altogether ca. 2100 pp.*

Abt's final report is composed of six separately bound documents:

Volume IVA: An Evaluation of Follow Through. 255 pp.

Volume IVB: Effects of Follow Through Models. 198 pp.

Volume IVC: Appendices. Part 1 and part 2, bound separately. Together 914 pp.

Volume IVD: Part I. A Longitudinal Study of Follow Through and Part II. Study of Follow Through Children at Academic Risk. 207 pp.

The Non-Follow Through Study. 568 pp.

These four volumes and the Non-Follow Through study are summarized separately.

Volume IVA: An Evaluation of Follow Through. This volume provides an overview of the analyses performed and the results they produced. Chapter I describes the origins and development of the planned variation concept, highlighting some of the difficulties encountered in conducting the Follow Through quasi-experiment, and identifying the respective cohorts, streams, and grade levels analyzed by this and previous Abt reports.

Chapter II discusses the categorization of sponsors along several dimensions--goals and objectives, nature of the classroom environment, and approach to home learning environment. The learning domains emphasized by various sponsors--basic skills, cognitive-conceptual skills, and socio-emotional development (or affective outcomes)--are utilized throughout the report in referring to types of effects. Chapter II also describes the instruments used to measure effects on children: Wide Range Achievement Test; Peabody Picture Vocabulary Test; Caldwell Preschool Inventory; Gumpgookies; Intellectual Achievement Responsibility Scale; Metropolitan Achievement Test-Primer, Primary I, Primary II and Elementary; Raven's Coloured Progressive Matrices; and Coopersmith Self-Esteem Inventory. Instruments used to measure outcomes in respective learning domains and at different grade levels are identified.

Chapter III describes the samples used in subsequent analyses--i.e., children from Cohort II (entering kindergarten and entering first grade) and Cohort III (entering kindergarten and entering first grade). The number of children from each of seventeen sponsors to be found in each of the four groups are presented. Basic demographic data of children in all four groups for respective sponsors are compared with each other and with the analytic sample as a whole.

---

*Abt's volumes IVA and IVB became volumes IIA and IIB in OE's final report The Follow Through Planned Variation Experiment. Also, two additional volumes in the Abt IV series not described here are scheduled for release in September 1977. These volumes, IVE and IVF in the Abt series, will become IIC and IIC appendix in the OE final report.

Chapter IV discusses the level of analysis used (site), the unit of analysis employed (child), efforts to adjust for noncomparability of groups; and methods for dealing with preschool influence. Three types of comparisons are used in analyzing student outcomes: local--a comparison of FT children with selected local NFT children within each site; pooled--FT outcomes at each site are compared with expected trends of performance as observed in the pooled sample of NFT children across all sponsors and sites; and normed--a comparison of FT and NFT performance at each site with national norms (for only the normed outcome measures). Two types of across-time comparisons are utilized: entry-exit for students with available data at entry kindergarten, and exit third grades and longitudinal for students with available data at entry kindergarten, end of first grade, and end of third grade. Adjustments for initial differences vary across analyses, but eleven background variables constitute the core adjustment variables. Preschool experience is also adjusted for, and analyzed as a separate factor, in some analyses. Also discussed are: the effects of overall attrition and attrition due to specific causes; results obtained using alternative sets of adjustment; the reliability of covariates; homogeneity of regression; and consequences of mismatch between Follow Through and comparison groups.

Chapter V presents and discusses the major findings, including: 1. The effectiveness of each FT model depends more on local circumstances than on the nature of the model. 2. Models that emphasize basic skills help children acquire these skills. 3. Unless a model emphasizes basic skills, children served by it are likely to score lower on tests of these skills than they would have done without Follow Through. 4. No single model or type of model is notably more successful than others in raising scores in cognitive conceptual skills. 5. Models emphasizing basic skills produce better results on tests of self-concept than other models do. 6. Some models are more successful in their most disadvantaged sites. 7. Direct Instruction (University of Oregon) and Bank Street are consistently more effective with Head Start children. 8. Most models are more effective during kindergarten and first grade than during second and third grade. 9. Some sponsors grew in effectiveness over time.

Volume IVB: Effects of Follow Through Models. The chapters of this volume are separately given to a discussion of seventeen sponsors. Each sponsor is described in terms of its particular pedagogical approach, its sample, and teachers' and parents' perceptions of the model. The comparability of each sponsor's sites and associated comparison groups are presented in three ways; the sponsors' FT sites are compared with all FT sites; the sponsors' comparison sites are compared to all NFT comparison sites; and the sponsors' FT sites are compared to the overall distribution of FT and NFT sites. In each case comparisons are drawn in terms of a measure of socioeconomic status and an ethnic-linguistic index. Also information is presented concerning teachers' and parents' perceptions of the model.

Estimates of effect on children resulting from a variety of analyses are then described. In each case effects estimates are presented in terms of tests of basic skills, cognitive conceptual measures and affective test outcomes. For most sponsors results are described from each of the following analyses: entry-exit contrasts using both local and pooled comparison; national norms comparisons; contrasts by type of preschool experience and longitudinal comparisons using a subsample of cohort III-EK children. Each sponsor chapter is closed with a summary.

Volume IVC: Appendices. The nine-hundred pages of this volume contain tables of data on matters ranging from distributions of family, child, teacher characteristics to treatment regression coefficients and standard errors by outcome. Results of the different comparisons and analytic strategies are presented by cohort-stream, sponsor and where available by site.

Volume IVD, Part I: A Longitudinal Study of Follow Through. The first part of this volume describes the overall analytic strategy and specific procedures used to examine sponsor effects on academic outcomes over two time periods--beginning K to end of first grade and end of first grade to end of third grade. The analysis is limited to Cohort III-EK children. Findings obtained in the analysis of each of eleven sponsors are presented. Overall conclusions are: (1) FT programs are relatively more effective in the quantitative than in the verbal skill area. (2) A variety of sponsors produce notable progress in the quantitative realm in the earlier years. However, verbal gains are more likely to be achieved by programs with a direct academic emphasis. (3) Head Start groups, more than other preschool groups, manifest a tendency to gain relatively less over the first two years.

Volume IVD, Part II: Study of Follow Through Children at Academic Risk. The second part of this volume describes an examination of longitudinal patterns of test scores for four groups of children taken from Cohort III-EK: FT-high risk, FT-low risk, NFT-high risk, and NFT-low risk children. High and low risk are operationally defined in terms of students' or their parents' scores on four measures--family income, educational level attained by head of household, children's scores on PPVT, and parents' scores on the locus-of-control scale. WRAT scores, entry and exit kindergarten, and standardized MAT scores from end of kindergarten, end of first grade, end of second grade, and end of third grade are pooled across all sponsors within each risk group and graphed for comparison. Sponsor-level means (and standard deviations) on each of six tests-time period combinations are reported. The findings of the study are that (1) For the majority of children enrolled Follow Through has not diminished the traditional relationship of greater disadvantage with lower school performance; (2) None of the individual sponsors materially diminished the relationship between disadvantaged status and poor academic test performance (although Direct Instruction children show small gains in reading and SEDL children show rather dramatic gains in mathematics); and (3) with some minor variations, patterns of achievement outcomes are quite similar for the four risk indicators.

Non-Follow Through Study. Focusing exclusively on NFT children, this separately bound report examines: (1) the relationships of background characteristics to academic performance at entry to kindergarten, (2) the relationship between background variables and school performance through the third grade, and (3) attrition in the (NFT) sample. Findings in each of these three areas of inquiry are discussed and supplemented with 445 pages of appendices, addenda, and appendices to addenda containing detailed information on the distribution of background characteristics and test scores obtained from the many instruments used at various time points.

Stephens, Mark S. Final Report. Evaluation of Locus of Control Measures for Follow Through Evaluation. Sponsored by BESE, OE/DHEW. Lafayette, IN: Undated (ca. 1971). ca. 50 pp.

The purpose of the study described in this report was "to pilot test three locus of control tests for use in evaluating the impact of different Follow Through program approaches." The three tests evaluated were Stephens-Delays Reinforcement Contengency Interview (SDRCI); the Gruen-Korte-Stephens (GKS) test; and the Nowicki-Strickland Locus of Control Scale for Chidren (NSLCSC). The study concluded that "It does appear that the SDRCI discriminates as expected between EDC and EB [Engelmann-Becker groups of children], and also between FT and NFT comparison groups. ...The NSLCSC had inadequate reliability and quite possibly because of this, no correlation with other measures,... The GKS was the only test that successfully predicted WRAT scores. However it did not discriminate between the NFT and FT classes, and it reflected differences opposite to those expected between EDC and EB groups."

Sullivan, Thomas F. "Health Care Delivery in Project Follow Through." Cambridge, MA: Bio-Dynamics, Inc., 1970. 30 pp.

A purposive sampling and study of 80 Follow Through sites yields information from 55 Follow Through projects on the successes and failures experienced in providing health services. Recommendations are made concerning Follow Through Health Care Component, Consultation Processes, and Alternative Procedures for providing technical Health Care Delivery.

Sutherland, Connie L., Middleton, Theresa, and Wright, Anne R. Spring 1975 Teaching Staff Questionnaire Cross Tabular Findings. Sponsored by OE/DHEW. Menlo Park, CA: Stanford Research Institute, May 1976.

"This report describes the findings from a cross-tabular analysis of the Spring 1975 Follow Through teacher questionnaire data" (p. 1). Findings concerning FT/NFT teacher differences, and respective sponsor deviations from NFT teachers in five areas are highlighted: educational goals and practices; adequacy of materials, staff, and training; parent participation and interactions; satisfaction - with student progress, with school and classroom environment, and with self; and miscellaneous.

Villaume, John and Haney, Walt. "Appendix to the Analysis of Interim Follow Through Evaluation Reports." [A summary and critical review of national Follow Through evaluation efforts.] Sponsored by OE/DHEW. Cambridge, MA: Huron Institute, 1976. 332 pp.

This is a draft version of the 1977 Project Follow Through National Evaluation Bibliography by the same authors. It differs from the latter study in two respects. First, it contains a much more detailed summary of the major national Follow Through Evaluation reports completed prior to 1976. Second, it contains a much more limited bibliography of literature on Follow Through.

Wax, Murray L. and Bruening, Robert G. Study of the Community Impact of the Hopi Follow Through Program. Final Report. Sponsored by the National Institute of Education/DHEW. Lawrence, KS: Kansas University Department of Sociology, 1973. 65 pp. ED 096 037.

This study examines how the implementation of a Follow Through project to serve children of the Hopi tribe in Arizona affected parents' definition of formal education and the school. Participant observation, semi-formal interviews, and a parent questionnaire were used to collect data during the period 1970 to 1972.

Weisberg, Herbert I. and Haney, Walt. Longitudinal Evaluation of Head Start Planned Variation and Follow Through. Cambridge, MA: Huron Institute, April 1977.

This study reports on analyses of data for children who participated in both Head Start Planned Variation and Follow Through. Using data from both Head Start and Follow Through years, the study focused on two main issues:

- Whether Follow Through helps maintain the benefits of Head Start in the early elementary grades;
- The way in which the Head Start experience of children may have confounded efforts in the national evaluation of Follow Through to calculate program effects.

The authors conclude "Follow Through does appear to maintain increased test scores on skills measured through the end of kindergarten. There are, however, no overall effects on skills measured in later grades. [Also,] we conclude that there is probably a bias operating against Follow Through in general, and particularly in sites where the Follow Through group has a much higher proportion of Head Start children than the Non-Follow Through group. This bias is not overwhelming, but may be large enough to change the outcomes significantly for many sites. In addition, we note that the design for data collection in Follow Through severely limits our ability to investigate how Follow Through has affected children's growth."

Wright, Anne R. Spring 1975 Parent Interview Cross-Tabular Findings. Sponsored by OE/DHEW. Menlo Park, CA: Stanford Research Institute, May 1976. 144 pp. Revised version: November 1976. 150 pp.

"This report describes the findings from the cross-tabular analysis of a subset of the Spring 1975 parent interview data" (p. 2), an analysis undertaken to identify discriminating information that could be used in a proposed Follow Through continuing effects study. Data collected from 3,183 FT and NFT parents related to 10 FT program models were collected. Sponsor deviations from FT averages and FT/NFT differences are reported on six dimensions: (1) parent perception of child's educational development, (2) child and parent attitudes toward school, (3) parent participation and interaction with school personnel, (4) child's home environment, (5) child's emotional adjustment, and (6) parent locus-of-control.

Wright, Anne R. Summary Report on Follow Through Testing Program 1969-1975. Menlo Park, CA: Stanford Research Institute, September 1976. 68 pp.

"This report describes the administration of pupil tests from the fall of 1969 through the spring of 1975 and the sample is described in terms of cohorts and entering level" (p. 1). Tests utilized in the national evaluation, data collection cycles, and scoring information are provided. In addition the report describes: (1) two pretests (Fall 1971 and Spring 1972) of unstandardized, non-norm-referenced tests that SRI was developing to assess non-cognitive student outcomes, (2) a City Study, (3) the Head Start Planned Variation Study, and the Summer Study.

III. OTHER DOCUMENTS

We include in this section a wide range of items, organized into the following nine categories:

1. Congressional Hearings and Reports, Items from the Federal Register, and GAO Reports.
2. Follow Through Evaluation Technical Papers and Feasibility Studies.
3. Follow Through-Related Conference Reports
4. Books and Articles
5. Theses
6. Papers Presented at Meetings and Conferences
7. Follow Through-Related Materials Produced by State Departments of Education or School Districts.
8. Monographs By Follow Through Sponsors or About Particular Sponsor Approaches
9. Miscellanea

These headings indicate a number of dimensions which were used to classify items--sponsorship, nature of the document, format in which the work appeared, and failure to meet the classification requirements of any other category. Some items would logically fall into more than one category--e.g., papers presented at a meeting or conference dealing with a particular sponsor's approach. In such cases, items are listed in the category which seems more obvious. In any case, items are included under only one category.

3.1 Congressional Hearings and Reports, Items from the Federal Register and GAO Reports.

Comptroller General of the U.S. Follow Through: Lessons Learned from Its Evaluation and Need to Improve Its Administration. Report to the Congress by the Comptroller General of the United States. Washington, D.C.: Comptroller General, 1975. 56 pp. ED 118 587.

Congress of the United States - House of Representatives

Congressional Record:
- Message on Children and Youth by President Lyndon B. Johnson, 90th Cong., 1st sess., February 8, 1967.
- Representative Danuts, December 12, 1969.
- Representative Dellenback, January 26, 1970.

Hearings before the Appropriations Committee and Subcommittees:
- 90th Cong., 2nd sess., Appropriations for FY 1969, Departments of Labor and Health, Education and Welfare, April, May and July, 1968.
- 91st Cong., 1st sess., Appropriations for FY 1970, Departments of Labor and Health, Education and Welfare, November, 1969.
- 91st Cong., 2nd sess., Appropriations for FY 1971, Departments of Labor and Health, Education and Welfare, May, 1970.
- 92nd Cong., 1st sess., Appropriations for FY 1972, Departments of Labor and Health, Education and Welfare, Part I, pp. 282 ff.
- 92nd Cong., 2nd sess., Appropriations for FY 1973, Departments of Labor and Health, Education and Welfare (Parts 2, 7), Spring 1972.
- 93rd Cong., 1st sess., Appropriations for FY 1974, Departments of Labor and Health, Education and Welfare, March, 1973.
- 93rd Cong., 2nd sess., Appropriations for FY 1975, Departments of Labor and Health, Education and Welfare, (Part 5), April and May, 1974.
- 94th Cong., 2nd sess., Appropriations for FY 1977, Departments of Labor and Health, Education and Welfare, February and March, 1976.

Hearings of the Committee on Education and Labor:
- 90th Cong., 1st sess. Lisle C. Carter, Jr., Assistant Secretary for Individual and Family Services, OE/DHEW. June 23, 1967.
- 90th Cong., 1st sess. Jule Sugarman, Associate Director, Head Start, Office of Economic Opportunity, June 26, 1967.
- 93rd Cong., 1st sess. Robert Harold, Principal, Portland, Oregon, February 21, 1973.
- 93rd Cong., 2nd sess. Sharel Hunt, Kindergarten Coordinator, Statewide Action on Follow Through, and Julia Balfour, Tacoma, Washington, February 1974.
  Recorded testimony accompanied by the following documents:

Wing Barefoot. "Learning Becomes a Family Venture," Tampa Times, Feb. 4, 1974.
Milton Goldberg. "Hopeful Signs for Urban Education."
Jane Stallings. "Is Compensatory Education Working?"
Robert L. Egbert. "Chapter 2 - Planned Variations in Follow Through."
Arkansas Department of Education. "[FT] Follow Through Report 1972-73."

Committee Print for the Committee on Education and Labor:
-92nd Cong., 1st sess. A Compilation of Federal Education Laws. USGPO, 1971.

Congress of the United States - Senate:

Congressional Record:
-Senator Clark, September 22, 29, 1967.
-Senator Pastore, September 5, 6, 1968.
-Senator Stevenson, April 5, 1972.
-Senator Percy, April 27, 1972.

Hearings before the Committee on Appropriations:
-92nd Cong., 1st sess. Appropriations for FY 1972. Office of Education, Special Institutions, and Related Agencies Appropriations. Part 2.
-93rd Cong., 2nd sess. Appropriations for FY 1975. May 1974.

Reports of the Committee on Labor and Public Welfare:
-90th Cong., 1st sess. On S.2388 - Economic Opportunity Act.

Office of Education/DHEW. "Grants to State Educational Agencies for Development of Successful Follow Through Program Models," Federal Register, Volume 34, No. 39 (February 26, 1972), 4105.

Office of Education/DHEW. "Follow Through Program Proposed Rules," Federal Register, Volume 39 (March 5, 1974), 8341.

Office of Education/DHEW. "Follow Through Program Rules and Regulations," Federal Register, Volume 39, No. 121 (June 21, 1974), 22342-22351.

Office of Education/DHEW. "Follow Through Rules and Regulations," Federal Register, Volume 40, No. 77 (April 21, 1975), 17711-17721.

Office of Education/DHEW. "Follow Through Program Proposed Rules [concerning grants for demonstration activities]," Federal Register, Volume 41, No. 231 (November 30, 1976).

3.2 Follow Through Evaluation Technical Papers and Feasibility Studies

Acland, Henry. "Report on Teacher Effects: Sample and Design." Sponsored by OE/DHEW. Cambridge, MA: Huron Institute, January 1974. 15 pp.

David, Jane. "Final Report on Sample Selection." Sponsored by OE/DHEW. Cambridge, MA: Huron Institute, June 30, 1974. Attachment B.

David, Jane. "Rationale for Selection of Sites and Grade Levels for Spring 1972 Follow Through Testing." Cambridge, MA: Huron Institute, January 20, 1972.

David, Jane. "Report on Design of the Follow Through Summer Effects Study." An unpublished report in G.L. McDaniels, "Follow Through Evaluation: Source Documents." 28 pp.

David, Jane. "The Role of All Sites in the Evaluation, including Projected Test Samples Through Spring 1976." A memorandum to Garry McDaniels, January 22, 1973, in G.L. McDaniels, "Follow Through Evaluation: Source Documents." 39 pp.

Emrick, John A. Results of Studies into the Reliability and Validity of Selective Cognitive and Non-cognitive Measurement Procedures Developed for the National Follow Through Evaluation. Menlo Park, CA: Stanford Research Institute, June 15, 1973. 80 pp.

Geiser, Kenneth R. A Feasibility Assessment of the Study of Follow Through Parent and Teacher Effects. Final Report. Sponsored by OE/DHEW. Cambridge, MA: Huron Institute, 1976. 121 pp.

Haney, Walt. Units of Analysis Issues in the Evaluation of Project Follow Through or There Must Be Heresy in This Some Place. Sponsored by OE/DHEW. Cambridge, MA: Huron Institute, 1974. 125 pp.

Howland, P.G. and Kaskowitz, D.H. "Report on Data Coding Reliability, School Year 1973/74 Data Collection Period." Menlo Park, CA: Stanford Research Institute, September 1974. 19 pp.

Judd, David E. "The Menucha Task Force Report: Recommendations on the Dissemination of Follow Through Materials." Sponsored by OE/DHEW. Portland, OR: Nero and Associates, 1973. 39 pp.

Marx, Thomas. "Problems in Longitudinal Analysis: A First Look at Longitudinal Versus Cross Section Studies." Sponsored by OE/DHEW. Cambridge, MA: Huron Institute, 1973. 19 pp. ED 103 440.

McNeil, Judy T. and Bond, James T. "Technical Comments in Response to Abt Follow Through Report." Ypsilanti, MI: High/Scope Educational Research Foundation, February 24, 1976.

Monaghan, Anne C. "A Consideration of the Feasibility of a National Study of the Impact of Follow Through in the Public School System." Sponsored by OE/DHEW. Cambridge, MA: Huron Institute, 1976. 39 pp.

Mosher, Frederick. Follow Through Instrumentation Study. Draft version. Sponsored by OE/DHEW. Cambridge, MA: Huron Institute, 1975. 115 pp.

Mosher, Frederick. "Progress Report on the Instrumentation Study." Sponsored by OE/DHEW. Cambridge, MA: Huron Institute, 1974. 35 pp.

Smith, M. S. "Head Start/Follow-Through Planned Variations Longitudinal Study." A memorandum to Lois-ellin Datta and Garry McDaniels, May 11, 1972, in G.L. McDaniels, "Follow Through Evaluation: Source Documents." 15 pp.

Stanford Research Institute. Follow Through Evaluation Data Bank File Description All Years Beginning 1972-73, Parts 1 & 2. Menlo Park, CA: Stanford Research Institute, revised January 1975.

Stanford Research Institute. Follow Through Evaluation Data Summary. Menlo Park, CA: Stanford Research Institute, April 1975.

Stanford Research Institute. Manual of Testing Procedures: Spring 1975. Sponsored by OE/DHEW. Menlo Park, CA: Stanford Research Institute, 1975. 146 pp. ED 111 839.

Stanford Research Institute. Outlines of Follow Through Sponsor and National Evaluation Activities: 1971-1972. Menlo Park, CA: Stanford Research Institute, July 1972. 100 pp.

Weisberg, Herbert I. and Bryk, Anthony S. "Cross Sectional/Longitudinal Studies of Follow Through." Sponsored by OE/DHEW. Cambridge, MA: Huron Institute, 1975.

3.3 Follow Through-Related Conference Reports

Abt Associates. "Follow Through Design Conference, July 19-20, 1976 Conference Handout." Cambridge, MA: Abt Associates, 1976.

Abt Associates, Inc. National Analytic Evaluation of Follow Through: Second Consultant Conference. Agenda and Support Materials. Cambridge, MA: Abt Associates, 9-11 June 1975. 400 pp.

Abt Associates. National Analytic Evaluation of Follow Through: Third Consultant Conference. Agenda and Support Materials. Cambridge, MA: Abt Associates, 1-3 October 1975. 311 pp.

Breivogel, William F., editor. Proceedings of Follow Through Conference (Gainesville, Florida, December 9-10, 1969). Gainesville, FL: Florida University, April 1970. 91 pp. ED 047 373.

High/Scope Educational Research Foundation. [Report of the] Working Conference on Evaluation of Follow Through Objectives Not Measured by Academic Tests. Report of Conference Activities and Recommendations. Ypsilanti, MI: High/Scope Educational Research Foundation, October 1971. 22 pp.

High/Scope Educational Research Foundation. [Report of the] Second Working Conference on Evaluation of Follow Through Objectives, Recommendations and Position Statements. Ypsilanti, MI: High/Scope Educational Research Foundation, 1972. 24 pp.

High/Scope Educational Research Foundation. [Report of the] Third Annual Conference on Evaluation of Follow Through Objectives. Working Papers. Ypsilanti, MI: High/Scope Educational Research Foundation, October 1973. 45 pp.

High/Scope Educational Research Foundation. Report of the Fourth Annual Follow Through Working Conference. Ypsilanti, MI: High/Scope Educational Research Foundation, October 1974. 61 pp.

High/Scope Educational Research Foundation. Report of the Fifth Annual Follow Through Sponsor Conference. Ypsilanti, MI: High/Scope Educational Research Foundation, October 1975. 51 pp.

Stanford Research Institute. Follow Through Meeting. Menlo Park, CA: Stanford Research Institute, March 11, 1971. 70 pp.

3.4 Books and Articles

Abelson, Willa D. and others. "Effects of a Four-Year Follow Through Program on Economically Disadvantaged Children," *Journal of Educational Psychology*, LXVI, 5 (October 1974), 756-771. EJ 118 479.

Anderson, S.B. "From Textbooks to Reality: Social Researchers Face the Facts of Life in the World of the Disadvantaged" in J. Hellmuth, editor. *Disadvantaged Child*. *Compensatory Education*: *A National Debate*. *Volume 3*. New York: Brunner-Mazel, 1970. 226-237.

Becker, Wesley C. "Some Effects of Direct Instruction Methods in Teaching Disadvantaged Children in Project Follow Through," Forthcoming in *The Proceedings of the International Symposium on Behavior Therapy* (Minneapolis, October 1972). NY: Appleton, Century Croft. 25 pp.

Bordon, Juliet P. and others. "Extended Positive Effects of a Comprehensive Head Start-Follow Through Program Sequence on Academic Performance of Rural Disadvantaged Students," *Journal of Negro Education*, XLIV, 2 (Spring 1975), 149-160. EJ 118 524.

Bordon, Juliet P. and others. "Head-Start Can Contribute to Reading," *Elementary English*, LI, 6 (September 1974), 874-876. EJ 104 089.

Butler, Annie L. *Current Research in Early Childhood Education*: *A Compilation and Analysis for Program Planners*. Washington, DC: American Association of Elementary, Kindergarten, and Nursery Educators, 1970. 203 pp. ED 058 927.

Clark, Alice T. "Follow-up on Follow Through." *Measurement and Evaluation in Guidance*, VII, 4 (January 1975), 234-238. EJ 110 591.

Cohen, David K. "The Value of Social Experiments," in Alice M. Rivlin and P. Michael Timpane, editors. *Planned Variation in Education*: *Should We Give Up or Try Harder*? Washington, DC: The Brookings Institution, 1975, 147-175.

Datta, Lois-ellin. "Design of the Head Start Planned Variation Experiment," in Alice M. Rivlin and P. Michael Timpane, editors. *Planned Variation in Education*: *Should We Give Up or Try Harder*? Washington, DC: The Brookings Institution, 1975, 79-99.

Eckerson, Louise O. "Following Through with Follow Through," *American Education*, IX, 10. 7 pp.

Education Daily. February 23, 1972; May 17, 1972; July 28, 1972; August 14, 1972; April 18, 1973; May 2, 1974; May 31, 1974; March 20, 1975; June 8, 1977; June 14, 1977; August 1, 1977; August 10, 1977.

Egbert, Robert L. "Follow Through," *National Elementary Principal*, LI, 1 (September 1971), 104-109. EJ 043 972.

Egbert, Robert L. "Follow Through: Fulfilling the Promise of Head Start," in J. Hellmuth, editor. *Disadvantaged Child*: *Head Start and Early Intervention*. Volume 2. NY: Brunner-Mazel, 1968. 573-580.

Elliott, Rogers and Vasta, Ross. "The Modeling of Sharing: Effects Associated with Vicarious Reinforcement, Symbolization, Age, and Generalization," *Journal of Experimental Psychology*, X, 1 (August 1970), 8-15. EJ 028-894.

Elmore, Richard F. "Design of the Follow Through Experiment," in Alice M. Rivlin and P. Michael Timpane, editors. *Planned Variation in Education*: *Should We Give Up or Try Harder*? Washington, DC: The Brookings Institution, 1975, 23-46.

Elmore, Richard F. "Follow Through Planned Variation," in W. Williams and R. Elmore. *Social Program Implementation*. NY: Academic Press, 1976, 101-123.

Enzmann, Arthur M. "Developing New Teaching Teams," *Childhood Education*, XLVII, 3 (December 1970), 131-134. EJ 029 169.

Evans, Ellis P. *Contemporary Influences in Early Childhood Education*. NY: Holt, Rinehart and Winston, 1971. 377 pp. ED 053 797.

Fallon, Berlie J., editor. *40 Innovative Programs in Early Childhood Education*. Belmont, CA: Fearon Publishers, 1973. 285 pp. ED 093 468.

Gabb, Sally. "Tupelo Follows Through." *Southern Education Report*, IV, 6, (January-February 1969), 22-28. EJ 001 776.

Glickstein, Howard A. "Federal Educational Programs and Minority Groups," *Journal of Negro Education*, XXXVIII, 3 (Summer 1969), 303-314. EJ 009 279.

Goldberg, Milton and others. "Project Follow Through: In Philadelphia," *Educational Leadership*, XXIX, 1 (October 1971), 20-23. EJ 043 885.

Gordon, Ira J. "An Instructional Theory Approach to the Analysis of Selected Early Childhood Programs," in National Society for the Study of Education. *The 71st Yearbook*: *Early Childhood Education*. Chicago: University of Chicago Press, 1972. 203-228.

Grannis, Joseph C. "Community, Competence, and Individuation: The Effects of Different Controls in Educational Environments," IRCD Bulletin, X, 2 (Spring 1975).

Greenwood, Gordon E., et al. "Some Promising Approaches to Parent Involvement," Theory into Practice, XI, 3 (June 1972), 183-189. EJ 063-912.

Griffin, Joan B. "Follow Through," School Health Review, V, 4 (July-August 1974), 25-27. EJ 106 660.

Guttman, Daniel and Barry, William. The Shadow Government. NY: Pantheon Books, 1976.

Holzman, Seymour and Boes, Shirley. Compensatory Education: What Works to Help Disadvantaged Pupils. Arlington, VA: National School Public Relations Association, 1973. 67 pp. ED 085 439.

Lavatelli, Celia S. "Models of Early Childhood Education: Questions But No Answers," Contemporary Psychology: A Journal of Reviews, XVII, 1 (January 1972), 14-15.

Love, Harold D. and Stallings, Sharon Gannaway. "A Comparison of Children Who Attended Project Head Start Not Having a Follow Through Program and Children Who Attended Project Head Start Having a Follow Through Program," Education, XCI, 1 (September/October 1970), 88-91. EJ 027 536.

Lukas, Carol van Deusen. "Problems in Implementing Head Start Planned Variation Methods," in Alice M. Rivlin and P. Michael Timpane, editors. Planned Variation in Education: Should We Give Up or Try Harder? Washington, DC: The Brookings Institution, 1975, 113-125.

Maccoby, E.E. and Zellner, M. Experiments in Primary Education. Aspects of Project Follow Through. NY: Harcourt, Brace and Jovanovich, 1971. 132 pp.

McDaniels, Garry L. "Evaluation Problems in Follow Through," in Alice M. Rivlin and P. Michael Timpane, editors. Planned Variation in Education: Should We Give Up or Try Harder? Washington, DC: The Brookings Institution, 1975, 47-60.

McDaniels, Garry L. "The Evaluation of Follow Through," Educational Researcher, IV, 1 (December 1975), 7-11. EJ 138 829.

McDonald, Jeanne. "Denver National Follow Through Conference," Voices (newsletter of the EDC Follow Through Program), December, 1972.

Miller, Louise B. and Dyer, Jean L. "Four Preschool Programs: Their Dimensions and Effects," Monographs of the Society for Research in Child Development, XL, 5-6 (October 1975). EJ 138 519.

Moorefield, Story. "Morgan Follows Through," American Education VI, 1 (January-February 1970), 31-33. EJ 015 167.

Moorefield. Story. "To Keep the Things We Love," American Education, VI, 7 (August-September 1970), 6-8. EJ 024 802.

New York Times. February 9, 1967; December 16, 1967; February 20, 1969; March 3, 1970; October 30, 1970; and December 13, 1970.

Packer, Athol B. and Cage, Bob N. "Changing Attitudes of Mothers Toward Themselves and Education," Theory into Practice, XI, 3 (June 1972), 196-201.

Pederson, Clara A. "New Day in North Dakota: Changing Teachers and Changing Schools," Childhood Education, XLVII, 5 (February 1971), 245-249. EJ 033 157.

Raffel, Jeffrey A. "Experiments as Reforms. An Analysis of the Implementation of an Educational Experiment," Urban Education, XI, 1 (April 1976), 59-78.

Rhine, W. Ray. "Strategies for Evaluating Follow Through," in R.M. Rippy, editor. Studies in Transitional Evaluation. Berkeley, CA: McCutchan Publishing Corporation, 1973. 157-180.

Rhine, W. Ray and Spencer, Leilani M. "Effects of Follow Through on School Fearfulness Among Black Children," Journal of Negro Education, XLIV, 4 (Fall 1975), 446-453. EJ 133 689.

Rivlin, Alice M. Systematic Thinking for Social Action. Washington, DC: The Brookings Institution, 1971.

Rivlin, Alice M. and Timpane, P. Michael, editors. Planned Variation in Education: Should We Give Up or Try Harder? Washington, DC: The Brookings Institution, 1975. ED 114 986.

Ross, Sylvia and Zimiles, Herbert. "The Differentiated Child Behavior Observation System," Instructional Science, V, 2 (October 1976), 325-342. EJ 145 201.

Shapiro, Edna. "Educational Evaluation: Rethinking the Criteria of Competence," School Review, LXXXI, 4 (August 1973), 523-549.

Smith, Marshall S. "Design Strategies for Experimental Studies," in Alice M. Rivlin and P. Michael Timpane, editors. Planned Variation in Education: Should We Give Up or Try Harder? Washington, DC: The Brookings Institution, 1975, 127-146.

Smith, Marshall S. "Evaluation Findings in Head Start Planned Variation," in Alice M. Rivlin and P. Michael Timpane, editors. Planned Variation in Education: Should We Give Up or Try Harder? Washington, DC: The Brookings Institution, 1975, 101-111.

Smithberg, Lorraine. "Some Reflections on Staff Development as an Interaction Process in Bank Street's Follow Through," *Journal of Research and Development in Education*, V, 3 (Spring 1972), 40-46.

Soar, Robert S. and Soar, Ruth M. "Early Childhood Education: An Experimental Analysis of Selected Follow Through Programs: An Example of a Process Approach to Evaluation," *National Society for the Study of Education Yearbook 1971* (part 2), 229-259. EJ 054 328.

Stallings, Jane A. "Implementations and Child Effects of Teaching Practices in Follow Through Classrooms," *Monographs of the Society for Research in Child Development*, XL, 7-8 (December 1975). 133 pp. EJ 138 518.

Strickler, Eve. "Letter from North Dakota: Inviting the People to See," *Urban Review*, VII, 4 (October 1974), 262-273. EJ 111 803.

Tamir, Pinchas. "Observation As a Tool for Evaluation of Implementation," *Studies in Educational Evaluation*, I, 2 (Summer 1975), 77-94. EJ 143 957.

Tavris, Carol. "Compensatory Education: The Glass Is Half Full," *Psychology Today*, (September 1976), 63-74.

Washington Post. June 20, 1977.

Weber, E. *Early Childhood Education: Perspectives on Change*. Worthington, OH: Charles A. Jones Publishing Company, 1970.

Weikart, David P. and Benet, Bernard A. "Model Design Problems in Follow Through," in Alice M. Rivlin and P. Michael Timpane, editors. *Planned Variation in Education: Should We Give Up or Try Harder*. Washington, DC: The Brookings Institution, 1975, 61-77.

Weikart, David P. and Benet, Bernard A. "Planned Variations from the Perspective of a Model Sponsor," in W. Williams and R. Elmore, editors. *Social Program Implementation*. NY: Academic Press, 1976, 125-148. Originally appeared in *Policy Analysis*, I, 3 (1975).

White, Burton L. "Making Sense Out of Our Educational Priorities." A preprint of an article to appear in *Compact Magazine*, 1973. Cambridge, MA: Harvard Graduate School of Education Laboratory of Human Development. 1973. 6 pp. ED 085 087.

Wilcox, Preston. "Changing Concepts of Community," Educational Leadership, XXIX, 8 (May 1972), 681-686.

Zimmerman, Barry J. and Bergan, John R. "Intellectual Operations in Teacher Question-Asking Behavior," Merrill-Palmer Quarterly of Behavior and Development, XVII, 1 (January 1971), 19-26. EJ 032 895.

3.5 Theses

Bennett, LaVerne Richardson. A Study of the Effects of a Two-Year Follow Through Program on the Academic Achievement of Second Grade Pupils. A doctoral thesis. Northwestern State University of Louisiana, 1971. 96 pp.

David, Jane Lisa. Summer Study: A Two-Part Investigation of the Impact of Exposure to Schooling on Achievement Growth. A doctoral thesis. Harvard Graduate School of Education, 1974. 168 pp.

Doyle, Michael Vincent. An Investigation and Evaluation of Speech Education in Pre-School and Early Elementary Programs for the Disadvantaged. A doctoral dissertation. East Lansing, MI: Michigan State University, 1969. 151 pp. ED 044 410

Elmore, Richard F. Follow Through: Decisionmaking in a Large-Scale Social Experiment. A doctoral dissertation. Cambridge, MA: Harvard Graduate School of Education, 1976. 398 pp.

Foote, Harold. An Assessment of Motivation Among Indian Students in the Todd County School System. A masters thesis. Aberdeen, SD: Northern State College, 1970. 51 pp. ED 086 394.

Kinard, Jesse Edward. The Effect of Parental Involvement on Achievement of First and Second Siblings Who Have Attended Head Start and Follow Through Programs. A doctoral dissertation. Florida State University College of Education, 1974. 81 pp.

Mahan, James Maruice. The Effects of Instruction by Teachers and Teacher Aides Upon the Performance of Pupils in a Direct Instructional Program. A doctoral dissertation. University of Illinois at Urbana-Champaign, 1971. 173 pp. ED 068 907.

McGill, Audrey Janet. The Effectiveness of the Use of Puppets in Oral Language Development of Culturally Disadvantaged First-Grade Children. A doctoral dissertation. North Texas State University, 1970. 124 pp.

Middendorf, Lorna Ann. Mothers as Mediators of Change in the Perceptual and Learning Abilities of Inner City Kindergarten Children. A doctoral thesis. Rutgers University, The State University of New Jersey, 1970. 164 pp.

Oldiges, Donald Allan. A Longitudinal Study of the Effects of Early Childhood Compensatory Education Programs in Dayton, Ohio. A doctoral thesis. University of Wisconsin, 1969. 291 pp.

O'Piela, Joan Marie. Identification of Predictor Variables of Success in First Grade Reading in Culturally Disadvantaged Inner-City Children Who Have Had a Preschool Experience. A doctoral thesis. Wayne State University, 1971.

Phillips, Clyde K., Jr. *A Comparative Study of the Effects of a Head Start Follow Through Program and a Kindergarten Program upon the Cognitive Abilities and Self-Concepts of Children from Low Socio-Economic Environments*. A doctoral thesis. The University of Oklahoma, 1971. 89 pp.

Roberts, Joseph B. *A Cognitive-developmental Study and Evaluation of the Pupil Role as Decision-maker*. A doctoral dissertation. Columbia University, 1974.

Russo, Joseph Alva. *Learning and Behavior Patterns of Follow Through Students Associated with Parent-Teacher Attitude Congruence*. A doctoral thesis. University of Utah, 1972. 160 pp.

Sanders, Frank J. *A Study of the Impact of the Chattanooga Public Schools' Head Start Follow-Through Program 1967-70*. A doctoral thesis. The University of Tennessee, 1970. 108 pp.

Taylor, Vera Cook. *An Evaluation of Three Compensatory Education Kindergarten Programs*. A doctoral dissertation. University of Southern California, 1970. 154 pp.

Thursby, Marilyn Pearcy. *Effects of Head Start and Follow Through on Dependency Striving, Dependency Conflict, and Autonomous Achievement Striving*. A doctoral thesis. Duke University, 1970. 232 pp.

Tocco, Thomas S. *A Mapping of Parent-Child Self-Concept Transmission in Florida Model Follow Through Participants*. A doctoral thesis. The University of Florida, 1970. 87 pp.

Van Strat, Georgena. *The Reading Problem in Urban Schools: Who Has It and What Has Been Done About It?* An Ed. D. dissertation. University of Massachusetts, 1975. 155 pp. ED 110 938.

3.6 Papers Presented at Meetings and Conferences

Anderson, Richard B. "Follow Through: Testing One Model of Evaluation and Several Models of Compensation." A paper presented at the Annual Meeting of the AERA (60th, San Francisco, 19-23 April 1976). 20 pp. ED 120 238.

Beers, C. David. "An Anthropological Approach to Documenting Educational Change: The Follow Through Implementation Study." A paper presented at the Annual Meeting of the Society for Applied Anthropology (St. Louis, March 1976). 40 pp.

Beers, C. David. "Principles of Implementation: The Lessons of Follow Through." Paper presented the Annual Meeting of the AERA (San Francisco, 19-23 April 1976). 38 pp. ED 124 283.

Also appears as a separate document: Portland, OR: Nero Associates, Inc. and Portland State University School of Education, 1976.

Bushell, Don, Jr. "Evaluating Classroom Progress." A paper presented at the Annual Meeting of the AERA (New Orleans, February 25-March 1, 1973). 3 pp. ED 076 652.

Clark, Woodrow and Beers, C. David. "Ethical Considerations in the Anthropological Evaluation of Educational Programs." Paper presented at the Annual Meeting of the AERA (60th, San Francisco, April 19-23, 1976). 34 pp. ED 129 844.

Conrad, Eva E. "Effects of Tutor Training, Achievement, and Experimentation on Process and Product Peer Tutoring Variables. Research Summary." A summary accompanying a presentation at Section 4.06 of the 1976 Annual Conference of the American Educational Research Association. 8 pp. ED 124 523.

Datta, Lois-ellin. "Parent Involvement in Early Childhood Education: A Perspective from the United States." A paper presented at the Centre for Educational Research Innovation Conference on Early Childhood Education (Paris, France, October 1973). 120 pp. ED 088 587.

Datta, Lois-ellin. "Planned Variation: An Evaluation of an Evaluative Research Study." A paper presented at the National Association of Young Children Conference (Atlanta, Georgia, November 15-16, 1972). 15 pp. ED 077 568.

Egbert, Robert L. "Planned Variation in Follow Through." A paper prepared for the Brookings Institution Conference on Social Experimentation. Washington, D.C.: April 1973. 27 pp.

Ferb, Thomas and others. "Comparing Curricular Models in Project Follow Through from Kindergarten to Third Grade." A paper presented at the Annual Meeting of the AERA (San Francisco, April 19-23, 1976). 49 pp. ED 125 775.

Fesler, Elizabeth and others. "Changes in Scholastic Achievement of the Disadvantaged Children Enrolled in Follow Through PEP-IPI Project." A paper presented at the Annual Meeting of the AERA (Washington, D.C., March 31-April 3, 1975). 10 pp. ED 108 750.

Goldberg, Milton. "Follow Through in Philadelphia." A paper presented at the Brookings Institution Panel on Social Experimentation Conference on Planned Variation Techniques (April 30-May 1, 1973). 32 pp. ED 127 016.

Gordon, Ira J. "An Early Intervention Project: A Longitudinal Look." A paper presented at the Merrill-Palmer Conference on Research and Teaching of Infant Development (Detroit, Michigan, February 8, 1973). 73 pp. ED 093 498.

Gordon, Ira J. "The Florida Parent Education Model." A paper prepared for the Atlanta Follow Through Workshop, October 10-12, 1968. 8 pp. ED 028 139.

Grannis, Joseph C. and Jackson, David E. "Setting Structure and the Problem of Match." A paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, Louisiana, February 25-March 1, 1973). 12 pp. ED 080 565.

Grannis, Joseph C. and others. "Setting Structure, Involvement, and Developmental Status as Learner: Elements of the Problems of the Match." A paper presented at the Annual Meeting of the AERA (New Orleans, February 25-March 1, 1973). 32 pp. ED 080 561.

Green, Dan S. and others. "The Behavior Analysis Follow Through Strategy: A Multifaceted Approach." A paper presented at the Annual Meeting of the AERA (60th, San Francisco, April 19-23, 1976). 34 pp. ED 126 125.

Green, Michael B. "On Involvement." A paper presented at the Annual Meeting of the AERA (New Orleans, February 25-March 1, 1973). 7 pp. ED 080 564.

Greenwood, Gordon E. and others. "A Study of Changes in Parents Employed as Paraprofessionals in a Home Intervention Follow Through Program." A paper presented at the Annual Meeting of the AERA (Chicago, April 15-19, 1974). 17 pp. ED 089 872.

Hodges, Walter L. "Process Evaluation in the Behavior Oriented Prescriptive Teaching Approach." A paper presented in Symposium at the Annual Meeting of the Mid-South Education Research Association (2nd, November 1973). 49 pp. ED 100 977.

Jackson, Donald A. and Bushell, Don, Jr. "Monitoring Staff Performance: A Plan for Quality Control in Project Follow Through." A paper presented at the Annual Meeting of the American Psychological Association (83rd, Chicago, August 30-September 3, 1975). 36 pp. ED 121 444.

Kaplan, Leonard and Tocco, T. Salvatore. "Impact of the Florida Model Follow Through Program as a Sponsoring Agency." A paper presented at the American Psychological Association Meeting (Miami, September, 1970). 8 pp. ED 045 563.

Kaskowitz, David and Stallings, Jane. "An Assessment of Program Implementation in Project Follow Through." A paper presented at the Annual Meeting of the AERA (Washington, D.C.: March 30-April 3, 1975). 27 pp. ED 106 349.

Kennedy, Mary and others. "Comparing Curricular Models in Project Follow Through from Kindergarten to Third Grade." A Symposium presented at the Annual Meeting of the AERA (San Francisco, April 1976). 19 pp.

Love, John M. "Assessment Using an Observation Procedure in the Cognitively Oriented Curriculum." A paper presented at the Annual Meeting of the AERA (59th, Chicago, April 1974). 14 pp. ED 097 355.

Marshall, Gloria. "Methods for Minimizing Attrition in Field Studies." A paper presented at the Annual Meeting of the AERA (60th, San Francisco, April 19-23, 1976). 11 pp. ED 128 412.

McDaniels, Garry L. "Attrition as an Analytical Programmatic Problem for Head Start Planned Variation and Follow Through." A paper presented at the Annual Meeting of the American Psychological Association (Honolulu, August 1972). 13 pp.

McNamara, Thomas. "Follow Through Evaluation in Philadelphia. The Moderate Realist Philosophical Position as the Base for Educational Evaluation." A paper presented at the Annual Meeting of the AERA (New Orleans, February 25-March 1, 1973). 28 pp. ED 074 134.

McNamara, Thomas. "Some Aspects of the Exposure-Achievement Relationships in Follow Through in Philadelphia." A paper presented at the Annual Meeting of the AERA (Chicago, April 1974). 18 pp. ED 093 908.

Needels,Margaret and Jane A. Stallings. "Classroom Process Related to Absence Rate." A paper presented to the Annual Meeting of the AERA (Washington, D.C., March 31-April 3, 1975). 23 pp. ED 110 199.

Nordt, Martha. "Pupil Role Interview Content Analysis Summary." A paper presented at the Annual Meeting of the AERA (New Orleans, February 25-March 1, 1973). 4 pp. ED 080 563.

Porter, Andrew C. "Some Design and Analysis Concerns for Quasi-Experiments Such as Follow Through." A paper presented at the American Psychological Association Meetings, 1972. 25 pp. ED 072 10

Proper, Elizabeth C. and St. Pierre, Robert G. "The Effects of Attrition on the National Evaluation of Follow Through." A paper presented at the Annual Meeting of the AERA (60th, San Francisco, April 19-23, 1976). 12 pp. ED 124 583.

Rentfrow, Robert K. and others. "Development of Situational Task Methodology for the Evaluation of Process Outcomes in the Open Classroom." A paper presented at the Annual Meeting of the AERA (New Orleans, February 25-March 1, 1973). 21 pp. ED 073 837.

Rhine, W. Ray. "Follow Through: A Model for the Utilization of Behavioral and Social Sciences in Solving Social Problems." A paper presented at the Annual Meeting of the American Psychological Association (Montreal, August 1973). 10 pp.

Rhine, W. Ray. "Toward More Effective Utilization of Behavioral and Social Scientists in School-Community Intervention Programs." A paper presented at the Annual Meeting of the AERA (Chicago, April 1974). 9 pp.

Roberts, Joseph B. "The Impact of Behavioral Modification Based Programs Versus Developmental Based Programs on Pupil Role Development." A paper presented at the Annual Meeting of the AERA (New Orleans, February 25-March 1, 1973). 8 pp. ED 080 562.

Rogers, Donald D. and Flynn, Donald L. "Cost Analysis of Follow Through Projects: A Description of Methodology." A paper presented at the Annual Meeting of the AERA (60th, San Francisco, April 19-23, 1976). 79 pp. ED 121 807.

Ross, Sylvia and Zimiles, Herbert. "Children's Interactions in Follow Through Classrooms: The DCB Observational System." A paper presented at the Annual Meeting of the AERA (Chicago, April 1974). 9 pp. ED 093 954.

Shapiro, Edna. "Examining Criteria for Evaluating Educational Programs." A paper presented at the biennial meeting of the Society for Research in Child Development (Philadelphia, March 29-April 1, 1973). 9 pp. ED 078 913.

Siegel, Martin A. and Rosenshine, Barak. "Teacher Behavior and Student Achievement in the Berieter-Engelmann Follow Through Program." A paper presented at the Annual Meeting of the AERA (New Orleans, February 25-March 1, 1973). 33 pp. ED 076 564.

Soar, Robert S. "A Measure of Teacher Classroom Management." A paper presented at the 79th Annual Convention of the American Psychological Association (Washington, D.C.: September 3-7, 1971). 9 pp. ED 057 890.

Stallings, Jane A. "An Implementation Study of Seven Follow Through Models for Education." A paper presented at the Annual Meeting of the AERA (59th, Chicago, April 1974). 32 pp. ED 093 971.

Stallings, Jane A. "Relationships Between Classroom Instructional Practices and Child Development." A paper presented at the Annual Meeting of the AERA (Washington, D.C.: March 31-April 3, 1975). 19 pp. ED 110 200.

Stallings, Jane A. "A Study of Implementation in Seven Follow Through Educational Models and How Instructional Processes Relate to Child Outcomes." A paper presented at the National Invitational Conference on Research on Teacher Effects: An Examination by Policy-Makers and Researchers (Austin, Texas, November 3-5, 1975). 35 pp. ED 128 340.

Stallings, Jane. "What Happens in the Follow Through Program? Implications for Child Growth and Development." A paper presented at the Annual Convention of the American Psychological Association (80th, Honolulu, September 2-8, 1972). 42 pp. ED 071 755.

Stallings, Jane A. "What Teachers Do Does Make a Difference - A Study of Seven Follow Through Educational Models." A paper presented to the Early Childhood Conference on Evaluation (Anaheim, CA, August 1974). 14 pp. ED 095 186.

Stallings, Jane A., Baker, Phillip, and Steinmitz, Gerald. "What Happens in the Follow Through Program: Implications for Child Growth and Development," A paper prepared for the American Psychological Association Convention (Honolulu, August 1972).

Stebbins, Linda B., Proper, Elizabeth C., and Ames, Nancy. "Problems in Measuring Program Implementation." A paper presented at the Annual Meeting of the AERA (60th, San Francisco, April 19-23, 1976). 10 pp.

Stephens, Mark W. "Dimensions of Locus of Control: Impact of Early Educational Experiences." A paper presented at the Annual Convention of the American Psychological Association (80th, Honolulu, September 6, 1972). 8 pp. ED 072 847.

St. Pierre, Robert G. and Proper, Elizabeth C. "Attrition: Identification and Exploration in the National Follow Through Evaluation." A paper presented at the Annual Meeting of the AERA (60th, San Francisco, April 19-23, 1976). 11 pp. ED 128 419.

Swift, Marshall. "The Application of a Behavioral Effectiveness Measure in a Longitudinal Study." A paper presented at the Annual Meeting of the AERA (New Orleans, February 25-March 1, 1973). 9 pp. ED 077 555.

Tocco, T. Salvatore and Bridges, Charles M. "Mother-Child Self-Concept Transmission in Florida Model Follow Through Participants." A paper presented at the Annual Meeting of the AERA (New York, 1971). 34 pp. ED 047 079.

White, William F. "Information Feedback Systems (IFS) and Educational Goals." A paper presented at Session I of Southeastern Invitational Conference on Measurement in Education (11th, Athens, Georgia, 8 December 1972). 14 pp. ED 075 463.

3. 7 Follow Through-Related Materials Produced by State Departments of Education or School Districts

Barr, Jo. A Resume of the Findings of the Follow Through Project Evaluation, 1969-70. Wichita, KA: Wichita Unified School District, August 1970. 52 pp. ED 048 392.

Berkeley Unified School District. "Head Start Follow Through Operational Grant, 1967-68." Berkeley, CA: Berkeley Unified School District, 1967. 44 pp. ED 017 595.

Bryan, Clifford and others. Fourth Year Results in Experiments in Early Education: A Comparative Assessment of Project Follow Through, Project Read and the Basal Reading Programs. Grand Rapids, MI: Center for Educational Studies, 1971. 251 pp. ED 077 873.

Cobb, Herman and Hunter, Earl. An Evaluation of the Nichols Avenue Follow-Through Programs. Final Evaluation Report. Sponsored by Oregon University. Washington, D.C.: District of Columbia Public Schools Department of Research and Evaluation, 1971. 68 pp. ED 068 158.

Crouthamel, Willard W. and others. Evaluation of Follow Through Program 197-71. Atlanta, GA: Atlanta Public Schools, September 1971. 67 pp. ED 059 173.

Goette, Margaret. Interdependent Learning in an Open Classroom Setting: Dean Rusk Elementary School, 1972-73. Atlanta, GA: Atlanta Public Schools Research and Development Report, Volume 7, No. 7, August 1973. 43 pp. ED 086 751.

Goette, Margaret and Watts, Larry. Interdependent Learning in a Traditional Classroom Setting: John Hope Elementary School, 1972-73. Atlanta, CA: Atlanta Public Schools Research and Development Report, Volume 7, Number 6, August 1973. 50 pp. ED 086 750.

Goette, Margaret and Watts, Larry. Multi-Uses of Compensatory Programs in Reinforcing the Curriculum of Two Inner-City Schools: English Avenue Elementary School; English Avenue Primary School; 1972-73. Atlanta, GA: Atlanta Public Schools Research and Development Report, Volume 7, No. 8, September 1973. 58 pp. ED 086 765.

Goldsboro City Schools. *Follow Through Program*. Goldsboro, NC: Goldsboro City Schools, 1974. 27 pp. ED 108 151.

Goodwin, Judith. *Relationship of the Follow Through Evaluation to Local Needs*. Philadelphia, PA: Philadelphia School District, 1972. 8 pp. ED 076 260.

Goodwin, Judith. *A Summary of Five Major Evaluation Reports on the Follow Through Program in Philadelphia in 1972-1973*. Philadelphia, PA: Philadelphia School District, 1973. 16 pp. ED 104 550.

Goodwin, Judith. *A Summary of Six Major Evaluation Reports on Follow Through in Philadelphia, 1973-74*. Philadelphia, PA: Philadelphia School District Office of Research and Evaluation, 1975. 25 pp. ED 118 629.

Halasa, Ofelia and Fleming, Margaret. *Follow Through Project, Health-Education-Welfare Fund (Fund Number 88-1). 1971-1972 Evaluation*. Sponsored by OE/DHEW. Cleveland, OH: Cleveland Public Schools Division of Research and Development, 1973. 193 pp. ED 084 002.

Hartford Connecticut Public School System. *Early Childhood Education Program: The Hartford Connecticut Public School System's Program of Open Informal Education in the Primary Grades Pre-K through 3.* Hartford, CT: Hartford Public Schools, 1972. 10 pp. ED 071 738.

Hartford Follow Through Report I. Fort Lauderdale, FL: Nova University Behavioral Sciences Center, November 1972. 200 pp. ED 077 598.

Illinois State Office of the Superintendent of Public Instruction and the Springfield Department of Exceptional Children. "Project Follow Through." Menlo Park, CA: Stanford Research Institute, 1973. 45 pp. ED 077 557.

Loreno, Patricia A. *Pupil Performance in the Atlanta, Georgia Elementary Schools*. Atlanta: GA. Atlanta Public Schools *Research and Development Report*, Volume 6, Numbers 11-14, December 1972. 156 pp. ED 075 561.

McNamara, Thomas and others. *An Evaluation of the Follow Through Program in Philadelphia, 1971-72*. *Volume I: Program Personnel's, and Parents' Perceptions of the Program*. *Volume II: Pupils' Achievement Characteristics*. Philadelphia, PA: Philadelphia School District Office of Research and Evaluation, November 1972. 109 pp. and 272 pp. ED 123 283 and ED 123 284.

Mulconry, Thomas R. and Allen, Raymond J. Jr. *An Evaluation of School Remedial Programs with Eligible Non-public Schools in Hartford, 1971-72. E.S.E.A.-S.A.D.C. Project*. Hartford, CT: Hartford City Board of Education. 52 pp. ED 067 430.

Ohio State Department of Education. "People Working Together: Follow Through in Ohio." Columbus, OH: Ohio State Department of Education, 1972. 36 pp. ED 073 827. [a pamphlet.]

Peper, John B. and McNamara, Thomas C. *A Longitudinal Study of Achievement, Transience, and Continuance in Follow Through, 1968-1972. An Individual Tracking Study*. Philadelphia, PA: School District of Philadelphia Office of Research and Evaluation, 1972. 169 pp.

Philadelphia School District, "The Comprehensive Early Childhood Education Network." Philadelphia, PA: Philadelphia School District, ca. 1972. 14 pp. ED 071 733.

Philadelphia School District. "Continuance and Transience among Teachers and Pupils in the Follow Through Program in Philadelphia, 1968-1972." Philadelphia, PA: Philadelphia School District Office of Research and Evaluation, 1972. 79 pp. ED 084 276.

Plankenhorn, Andrew and Sylvan, Donna L. *The Interdependent Learning Model and Achievement: Grant Park Primary School, 1972-73. Research and Development Report*, Vol. 7, No. 39. Atlanta, GA: Atlanta Public Schools, 1974.

Plunkett, Virginia R.L. "Spotlight on Follow Through." Sponsored by OE/DHEW. Denver, CO: Colorado State Department of Education, 1969. 28 pp. ED 029 720.

Prince George's County Board of Education. *Evaluation of Follow Through, 1968-1969*. Upper Marlboro, MD: Prince George's County Board of Education, 1969. 57 pp. ED 044 172.

Puzzoli, David A. and Fazzoro, Charles J. *Project ERA: A Three-Year Study of a Follow Through Program. A Longitudinal Study of the Monongolia County Follow Through Program*. Morgantown, WV: Monongolia County Board of Education, 1970. 123 pp. ED 048 948.

Richmond Public Schools. Home Learning Activities Designed to Provide Educational Experiences for Children and Parents: A Home Preschool Program Designed for the Siblings of the Children on the Follow Through Program in the Richmond Public Schools. Richmond, VA: Richmond Public Schools, ndg. 221 pp. ED 105 988.

Russell, Gail and Gaskin, Comer. "The Use of Several Diverse Methodologies for Increasing Pupil Achievement: Wesley Avenue Elementary School, 1972-73." Atlanta, GA: Atlanta Public Schools Research and Development Report, Volume 7, No. 30, November 1973. 45 pp. ED 090 342.

Seay, Shirley H., and Brown, Brenda. Home Learning Activities Designed to Provide Educational Experiences for Children and Parents. Parts 1 and 2. Richmond, VA: Richmond Public Schools, ndg. 200 pp. and 206 pp. ED 105 989 and ED 105 990.

Sweeney, David L. "Title I, E.S.E.A., 1967-1968. Revised Report." Racine, WI: Racine Unified School District, Number 1, 1967. 35 pp. ED 016 729.

Watson, Jeanette and Lanham, Fritz. Early Childhood Development in Texas. Special Report. Austin, TX: Texas State Department of Community Affairs Office of Early Childhood Development, 1971. 144 pp. ED 067 145.

Wichita Unified School District 259. Follow Through Project, Wichita Unified School District 259: Initial Year, September 1968-May 1969 Evaluation Report. Wichita, KS: Wichita Unified School District 259, 1969. 156 pp. ED 039 027.

3.8 Monographs by Follow Through Sponsors or About Particular Follow Through Sponsor Approaches*

American Institutes for the Study of the Behavioral Sciences. "Behavior Analysis Model of a Follow Through Program, Oraibi, Arizona; Childhood Education Model Programs." Sponsored by the National Center for Educational Communication, OE/DHEW, and the Office of Economic Opportunity. Palo Alto, CA: American Institutes for Research in the Behavioral Sciences, 1970. 24 pp. ED 045 251.

American Institutes for the Study of the Behavioral Sciences. "Behavior Principles Structural Model of a Follow Through Program, Dayton, Ohio: Model Programs. Childhood Education." Sponsored by the National Center for Educational Communications, OE/DHEW, and the Office of Economic Opportunity. Palo Alto, CA: American Institutes for Research in the Behavioral Sciences, 1970. 16 pp. ED 045 783.

American Institutes for the Study of the Behavioral Sciences. "Model Programs: Childhood Education. Interdependent Learner Model of a Follow Through Program." Sponsored by OE/DHEW and the Office of Economic Opportunity. Silver Spring, MD: American Institutes for Research in the Behavioral Sciences, 1970. 19 pp. ED 045 328.

American Institutes for the Study of the Behavioral Sciences. "Model Programs: Childhood Education. Responsive Environment Model of a Follow Through Program. Goldsboro, North Carolina." Sponsored by the National Center for Educational Communication, OE/DHEW, and the Office of Economic Opportunity. Palo Alto, CA: American Institutes for Research in the Behavioral Sciences, 1970. 26 pp. ED 046 108.

American Institutes for Research in the Behavioral Sciences. "Tacoma Public Schools Early Childhood Program, Tacoma, Washington: Combined Local, State, and Federal Funds Support a Large-Scale Early Childhood Program in the Public Schools. Model Programs-Childhood Education." Sponsored by the National Center for Educational Communication, OE/DHEW, and the Office of Economic Opportunity. Palo Alto, CA: American Institutes for Research in the Behavioral Sciences, 1970. 18 pp. ED 045 221.

Armington, David E. "A Plan for Continuing Growth." Newton, MA: Educational Development Center, 1969. 22 pp. ED 046 493.

---

*This listing of sponsor materials includes only documents identified in an April 1977 search of the ERIC system or used directly in preparing the technical history of the Follow Through evaluation. For more complete listing of sponsor materials, see Volume III of OE's final report on the Follow Through experiment, or the report by Judd and Wood, described in section II of this appendix.

Bank Street College of Education. "Report of Follow Through Parent Interview Project." New York: Bank Street College of Education, 1973. 36 pp. ED 086 362.

Becker, Wesley C. and Engelmann, Siegfried. "The Direct Instruction Model." Eugene, OR: University of Oregon, May 14, 1976, paper prepared for the USOE volume of sponsor chapters.

Becker, Wesley C. and Engelmann, Siegfried. "Program Description and 1973 Outcome Data: Engelmann-Becker Follow Through Model." Eugene, OR: Oregon University Department of Special Education, 1973. 45 pp. ED 096 780.

Becker, Wesley C. and Englemann, Siegfried. *Summary Analysis of Five-Year Data on Achievement and Teaching Progress with 14,000 Children in 20 Projects*. Technical Report 73-2. Preliminary Report. Eugene, OR: Oregon University Department of Special Education, 1973. 54 pp. ED 096 781.

Bowman, Garda W. and others. "Program Analysis System in Bank Street Follow Through." New York: Bank Street College of Education, April 1976. 27 pp.

Breivogel, W.F. and others. *Training of Teachers, Parent Educators, PAC, Administrators, in Florida Parent Education Follow Through Model*. *Final Report*. Sponsored by OE/DHEW. Gainesville, FL: Florida University Institute for the Development of Human Resources, 1973. 295 pp. ED 092 218.

Brigham, Thomas A. *Towards an Operant Analysis of the Acquisition of Conceptual Behavior*. Sponsored by OE/DHEW. Lawrence, KS: Kansas University Department of Human Development, ca. 1971. 53 pp. ED 060 935.

Connecticut University Educational Resources and Development Center. *A Study of the Bank Street Approach at the John J. Ryle School, Stamford, Connecticut.* Storrs, CT: Educational Resources and Development Center, 1975. 111 pp. ED 126 202.

Conrad, Eva E. *The Effects of Tutor Achievement Level, Reinforcement Training, and Expectancy on Peer Tutoring*. Sponsored by OE/DHEW. Tucson, AZ: Arizona Center for Educational Research and Development, 1975. 76 pp. ED 116 807.

Education Development Center. "EDC Open Education Follow Through Program" August 1976, paper prepared for the USOE volume of Follow Through sponsor chapters.

Education Development Center. Instructional Aids, Materials, and Supplies--Guidelines. Newton, MA: Educational Development Center, 1971, 61 pp. ED 093 848.

Far West Laboratory for Educational Research and Development. "Florida Parent Education Program. Program Summary." Sponsored by OF/DHEW. Gainesville, FL: Florida University Institute for Development of Human Resources, 1971. 6 pp. ED 125 750.

Far West Laboratory for Educational Research and Development. The Interdependent Learning Model. Program Summary. New York: New York University Institute for Developmental Studies, 1971. 8 pp. ED 125 745.

Gordon, Edmund W. and others. "Ira Gordon Follow Through (Richmond, Virginia)." Sponsored by the National Center for Educational Communication, OE/DHEW. New York: Columbia University, 1972. 34 pp.

Gordon, Edmund W. and others. "Weikart Follow Through (Trinidad, Colorado)." Sponsored by the National Center for Educational Communication, OE/DHEW. New York: Columbia University, 1972. 48 pp. ED 126 180.

Gordon, Ira J. "Institute for Training Follow Through Personnel in Parent Participation and Education." Sponsored by OE/DHEW. Gainesville, FL: Florida University College of Education, 1968. 12 pp. ED 028 140.

Gordon, Ira J. and Guinagh, B.J. "A Home Learning Center Approach to Early Stimulation." Gainesville, FL: University of Florida Institute of Development of Human Resources, 1974. ca. 90 pp.

Gordon, Ira J. and others. Reaching the Child Through Parent Education. The Florida Approach. Gainesville, FL: Florida University for Development of Human Resources, February 1969. 130 pp. ED 057 880.

Greenwood, Gordon E. Teacher Morale in a Home Intervention Follow Through Program. 1970. 17 pp. ED 104 989.

Greenwood, Gordon E. Assistance to Local Follow Through Programs. Annual Report. Sponsored by OE/DHEW. Gainesville, FL: Florida University Institute for Development of Human Resources, 1972. 308 pp. ED 073 830.

Henderson, Ronald W. "Research and Consultation in the Natural Environment." Sponsored by OE/DHEW. Tuscon, AZ: Arizona Center for Early Childhood Education, 1969. 15 pp. ED 037 240.

Henderson, Ronald W. and others. "Arizona Center for Early Childhood Education. Final Report." Sponsored by OE/DHEW. St. Ann, MO: National Coordination Center for Early Childhood Education, 1972. 46 pp. ED 128 094.

High/Scope Educational Research Foundation. *Curriculum Development and Field Issues*. *Volume I, Progress Report, Cognitively Oriented Curriculum*, *Project Follow Through*. Sponsored by OE/DHEW. Ypsilanti, MI: High/Scope Educational Research Foundation, 1973. 196 pp. ED 088 579.

High/Scope Educational Research Foundation. *Sponsor Evaluation Activities*: *Formative and Summative*. *Volume II*. *Progress Report*, *Cognitively Oriented Curriculum*, *Project Follow Through*. Ypsilanti, MI: High/Scope Educational Research Foundation, 1973. 275 pp. ED 088 580.

High/Scope Educational Research Foundation. "Summary of Findings from the 1974-75 Assessment of Second and Third Grade Follow Through and Non-Follow Through Children's Written Language Using the High Scope Productive Language Arts Test." A draft report. Ypsilanti, MI: High/Scope Educational Research Foundation, 1975. 34 pp.

Hilliard, Grace R. *Final Report of the 1975 Survey of Adult EDC Follow Through Opinion*. Sponsored by OE/DHEW. Newton, MA: Education Development Center, January 30, 1976. 322 pp.

Hodges, Walter L. *The Implications of Design and Model Selection for the Evaluation of Programs for the Disadvantaged Child*. Conway, AR: Arkansas State College, December 1970. 24 pp. ED 047 792.

Jester, R. Emile. "Focus on Parent Education as a Means of Altering the Child's Environment." Sponsored by College of Education, Florida University, Gainesville and OE/DHEW. Gainesville, FL: Florida University College of Education, 1969. 11 pp. ED 033 758.

Love, John M. and Bond, James T. "The High/Scope Productive Language Arts Test in the Cognitive Curriculum Effects of Follow Through in the Written Language of Second and Third Grade Children." Sponsored by OE/DHEW. Ypsilanti, MI: High/Scope, March 1975. 50 pp.

Marion, David J., Nameche, Gene, and Rogers, Jacqueline. *Phase One 1975-76 Documentation Study*. Newton, MA: Education Development Center, December 1976.

Morris, Mary E. and Love, John M. *Classroom Interactions in Four Follow Through Sites*. *Volume III*, *Progress Report*, *Cognitively Oriented Curriculum*, *Project Follow Through*. Ypsilanti, MI: High Scope Educational Research Foundation, 1973. 354 pp. ED 088 581.

Nimnicht, Glen P. "Environmentally Deprived Children." July 1969. 10 pp. ED 039 937.

Nimnicht, Glen P. "Overview of Responsive Model Program." Sponsored by the National Center for Educational Research and Development, OE/DHEW. Berkeley, CA: Far West Laboratories for Educational Research and Development, 1970. 12 pp. ED 045 207.

Nimnicht, Glen P. "A Revision of the Basic Program Plan of Education at Age Three." Sponsored by the National Center for Educational Research and Development, OE/DHEW. Berkeley, CA: Far West Laboratories for Educational Research and Development, 1970. 37 pp. ED 047 774.

Nimnicht, Glen P. and others. "Preliminary Analysis of 1968-69 Booth Achievement." Sponsored by the National Center for Educational Research and Development, OE/DHEW. Berkeley, CA: Far West Laboratories for Educational Research and Development, 1969. 10 pp. ED 045 201

Ng, Pierina. "The Assignment of Children to Responsive Follow Through Classrooms," Undated draft. Berkeley, CA: Far West Laboratory for Educational Research and Development. 9 pp.

Rayder, Nicholas F. and others. *Implementation of Responsive Program: A Report on Four Planned Variation Communities*. Sponsored by the Office of Child Development/DHEW. Berkeley, CA: Far West Laboratory for Educational Research and Development, 1973. 200 pp. ED 085 102.

Rayder, Nicholas F. and others. *Preliminary Analysis on Kindergarten and First Grade Follow Through Test Results for 1968-69*. Sponsored by the National Center for Educational Research and Development, OE/DHEW. Berkeley, CA: Far West Laboratories for Education Research and Development, 1970. 26 pp. ED 045 202.

Rayder, Nicholas F. and others. *The Responsive Classroom Observation Schedule--Background and Development*. San Francisco, CA: Far West Laboratories for Educational Research and Development, 1975. 39 pp. ED 107 375.

University of North Dakota Center for Teaching and Learning. "The New School Approach." 1976, paper prepared for the USOE volume of Follow Through sponsor chapters. 20 pp.

Ware, William B. and others. *Assistance to Local Follow Through Programs*. *Annual Report*. Sponsored by OE/DHEW. Gainesville, FL: Florida University Institute for Development of Human Resources, 1973. 274 pp. ED 094 852.

Ware, William B. and others. *Assistance to Local Follow Through Programs*. *Annual Report*. Sponsored by OE/DHEW. Gainesville, FL: Florida University Institute for Development of Human Resources, 1974. 296 pp. ED 107 349.

Weikart, David. "Alternative Assessment and National Follow Through," *High/Scope Educational Research Foundation Report 1975-76*. Ypsilanti, MI: High/Scope Educational Research Foundation, 1976.

Weiss, Carol N.; Hamilton, Stephen F.; Hirano, Marsha J. *An Assessment of the Impact of the EDC Follow Through Advisory Approach*. A Study based on interviews with teachers, aides, administrators and parents in the EDC Follow Through Program in Burlington, Vermont, 1973-1974. Sponsored by OE/DHEW. Newton, MA: Education Development Center, 1974. 253 pp.

Weiss, Carol N.; Hamilton, Stephen F.; and Hirano, Marsha J. *An Assessment of the Impact of the EDC Follow Through Advisory Approach*. A study based on interviews with teachers, aides, administrators and parents in the EDC Follow Through Program in Paterson, New Jersey, 1973-1974. Sponsored by OE/DHEW. Newton, MA: Education Development Center, 1974. 231 pp. ED 110 424.

Zimmerman, Barry J. and Bergan, John R. "Intellectual Operations in Teacher-Child Interaction." Sponsored by OE/DHEW. Tucson, AZ: University of Arizona, 1968. 16 pp. ED 039 011.

3.9 Miscellanea

Abt Associates. "A Proposal to Conduct an Analytic Study of the Follow Through Program 1969-72. Part I: Technical Proposal." RFP No. 72-37, AAI #1507, May 5, 1972. Cambridge, MA: Abt Associates, 1972.

Abt Associates. "Analytic Evaluation of the Follow Through Experiment. Interim Report I," 21 March 1975. Cambridge, MA: Abt Associates, 1975.

Alkin, Marvin. "A Review of the Evaluation of the Follow Through Program." Sponsored by OE/DHEW. Bureau of Research, Los Angeles, CA: UCLA Center for the Study of Evaluation, 1970. 9 pp. ED 941 462.

Ames, Nancy. Memorandum "Implementation Study" to Richard B. Anderson. Cambridge, MA: Abt Associates, undated [1975].

Ames, Nancy, Anderson, R.B., and Cline, M. et al. "A Proposal to Analyze Selective Sets of Data Collected on the Follow Through Program. RFP HEW/OE 75-2. Responses to Technical Questions." Cambridge, MA: Abt Associates, December 23, 1974.

Anderson, Richard B. Memorandum "Implementation and Comparability" to "People Who Know About the Realities of Follow Through Sites. Cambridge, MA: Abt Associates, March 19, 1975.

Ballentine, Larry and Levine, Daniel U. "Research Note: Home Environment and Reading Performance among Afro, Anglo, and Mexican Kindergarten Students in an Inner City School." Kansas City, MO: Missouri University Center for the Study of Metropolitan Problems in Education. 8 pp. ED 058 943.

Bank Street College of Education. Follow Through Task Force Reports. New York: Bank Street College of Education, undated [1967]. ca. 55 pp.

Bissell, Joan S. Implementation of Planned Variation in Head Start. I. Review and Summary of the Stanford Research Institute Interim Report: First Year of Evaluation. Bethesda, MD: National Institute of Child Health and Human Development, 1971. 51 pp. ED 052 845.

Bissell, Joan S. Planned Variation in Head Start and Follow Through. Washington, D.C.: DHEW, 1972. 72 pp. ED 069 355.

Bronfenbrenner, Urie. "Motivational and Social Components in Compensatory Education Programs: Suggested Principles, Practices, and Research Designs." Washington, D.C.: OE/DHEW, 1968. 34 pp. ED 024 464.

Crooks, J. and Burnes, D. "Consolidated Review of September 17 Deliverables," memorandum to F. Bresnick, Acting Chief, Research and Evaluation Section, Follow Through Program, October 6, 1971.

Deloria, Vine, Jr., editor. Indian Education Confronts the Seventies. [Five Volumes.] Volume II: Theoretical Considerations in Indian Education. Sponsored by OE/DHEW, Office of Indian Education. Oglada, SD: American Indian Resource Associates and Tsaile, AZ: Navajo Community College, 1974. 260 pp. ED 113 067.

Department of Health, Education, and Welfare, Audit Agency. Draft Report on Review of Administration of and Performance Under Contracts with the Stanford Research Institute. Washington, D.C.: OE and Office of Child Development, December 27, 1972. 50 pp.

DHEW. "Follow Through: Promising Approaches to Early Childhood Education." Washington, D.C.: U.S. Government Printing Office, 1971. 26 pp.

DuPree, T.J. "Brief History of Cherokee Schools 1804-1976." Albuquerque, NM: Bureau of Indian Affairs, May 1976. 13 pp. ED 127 051.

Durost, W. N. et al. Metropolitan Achievement Test Tester's Directions. New York: Harcourt Brace, 1971

Edward J. Cherian and Associates, Inc. A Guide to Follow Through. Washington, D.C.: OE/DHEW, 1973. 117 pp.

Egbert, Robert L. "Follow Through." An unpublished paper. 66 pp.

Foster, Florence P. Planning Parent-Implemented Programs: A Guide for Parents, Schools and Communities. Washington, D.C.: OE/DHEW, 6 February 1969. 53 pp. ED 036 322.

Gilkeson, Elizabeth. Letter to Director Follow Through Evaluation, Abt Associates, February 18, 1976, enclosure. New York: Bank Street College of Education, 1976.

Glockner, Mary. "Worktable on Wheels." Sponsored by the Office of Economic Opportunity. Urbana, IL: ERIC Clearinghouse on Early Childhood Education, 1969. 2 pp. ED 035 451.

Gordon, Edmund W. and others. "A New Lease on Education Through Health and Nutrition Programs (San Diego, Texas)." Sponsored by the National Center for Educational Communication, OE/DHEW. New York: Columbia University, 1972. 38 pp. ED 126 181.

Grannis, Joseph C. "Task Engagement and the Congruence of Pedagogical Controls: An Observation Study of Differently Structured Classrooms." An unpublished paper circulated at Teachers College, Columbia University.

Hilliard, Grace R. "Overall Goals and Objectives of the EDC Follow Through Program," letter to Mary Kennedy, February 6, 1976. Newton, MA: Education Development Center, 1976.

Horton, Robert and Bryan, Clifford. *Accountability and Teacher Satisfaction*. Kalamazoo, MI: Western Michigan University, Department of Sociology, 1971. 11 pp. ED 063 270.

Hymes, James L., Jr. *Early Childhood Education: An Introduction to the Profession*. Washington, D.C.: National Association for the Education of Young Children, 1968. 48 pp. ED 036 328.

Jablonsky, Adelaide, compiler. *Early Childhood Education for the Disadvantaged: An Annotated Bibliography of Doctoral Dissertations. ERIC-IRCD Doctoral Research Studies, Number 3, July 1973*. Sponsored by NIE/DHEW. New York: ERIC Clearinghouse on the Disadvantaged, 1973. 203 pp. ED 079 438.

Johnson, Edward E. Letter to Dr. Mary M. Kennedy, February 20, 1976.

Kennedy, Mary. "Conversations with Follow Through Sponsors Staff." An unpublished transcription of recordings made December 1975. ca. 45 pp.

Klein, Jenny. "Head Start Planned Variation Program." September 1969. 5 pp. ED 038 170.

Lazar, Joyce B. and Chapman, Judith E. *A Review of the Present Status and Future Research Needs of Programs to Develop Parenting Skills*. Washington, D.C.: George Washington University Social Research Group, 1972. 206 pp. ED 068 150.

Lutsky, Judi, compiler. *Head Start and Follow Through, 1972-74: An ERIC Abstract Bibliography*. Sponsored by the National Institute of Education/DHEW. Urbana, IL: ERIC Clearinghouse on Early Childhood Education, 1974. 61 pp. ED 097 131.

Masla, John A. and Roysten, Preston M. *Community Involvement in Teacher Education: A Study of the Models*. Sponsored by the National Institute of Education/DHEW. Washington, D.C.: ERIC Clearinghouse on Teacher Education, 1976. 93 pp. ED 121 708.

McDaniels, Garry L. "Follow Through Evaluation: Source Documents." [A compendium of eight documents, many untitled, having to do with Follow Through evaluation.]

Mergendoller, John R. Compensatory Education and the Varieties of Intervention: A Program Element Typology Based on the Model of the Follow Through Planned Variation Experiment. Washington, D.C.: National Institute of Education, 1976. 109 pp. ED 129 443.

Miller, Louise B. and others. "Experimental Variation of Head Start Curricula: A Comparison of Current Approaches. Progress Report No. 9, March 1, 1971-May 31, 1971." Sponsored by the Office of Economic Opportunity. Louisville, KY: Louisville University Department of Psychology, 1971. 26 pp. ED 053 814.

National Advisory Council on the Education of Disadvantaged Children. "Educating the Disadvantaged Child: Where We Stand. The 1972 Annual Report to the President and the Congress." Washington, D.C.: NACEDC, 1972. 49 pp. ED 065 649.

Office of Child Development, DHEW. "Head Start Planned Variation Study." Washington, D.C.: Office of Child Development, 1970. 17 pp. ED 047 782.

Office of Education/DHEW. Annual Evaluation Report on Programs Administered by the U.S. Office of Education FY 1975, Washington, D.C.: OE, OPBE, undated [1976].

Office of Education/DHEW. "Draft Follow Through Program Manual." Washington, D.C.: OE, February 24, 1969.

Office of Education/DHEW. "Follow Through: Promising Approaches to Early Childhood Education, School Year 1970-71." Washington, D.C.: OPBE, USOE, September 1974.

Office of Education. "Request for Proposal. Analysis of Follow Through Evaluation Data." OPBE, USOE, September 1974.

Office of Education. "Request for Proposal, Follow Through, Draft." Washington, D.C.: OE, March 13, 1972.

Office of Education/DHEW. Second Report of the Follow Through Evaluation Review Panel. March 1973. 56 pp.

Office of Education. "Supporting Statement For OMB Clearance of the Spring 1972 Parent Interview Instrument, January 4, 1972."

Quilliam, Benjamin F., Jr. and Rogers, Kathryn S. Nine Model Programs for Young Children. Volumes I and II. Sponsored by National Institute of Education/DHEW. St. Ann, MO: National Coordination Center for Early Childhood Education, 1972. 184 pp. and 137 pp. ED 129 402 and ED 129 403.

RMC. RMC Proposal No. 279. Cost Analysis Study of Follow Through Projects. Bethesda, MD: RMC Research Corporation, April 18, 1975. ca. 400 pp.

RMC. "Proposal No. 279. Cost Analysis Study of Follow Through Projects Technical Amendment." Bethesda, MD: RMC Research Corporation, May 12, 1975. 22 pp.

RMC. "Proposal No. 279. Cost Analysis of Follow Through Projects. Costing Methodology." Bethesda, MD: RMC Research Corporation, September 10, 1975. ca. 35 pp.

Rodriguez, Norma. Modular Sequence: Teaching Reading to Bilingual Learners. Hartford, CT: Hartford University College of Education, ndg. 35 pp. ED 106 251.

Salzer, Richard T. Care and Education of the Young Child. Sponsored by OE/DHEW. Denver, CO: National Association of State Boards of Education, 1972. 19 pp. ED 071 758.

Shipman, Virginia, project director. Disadvantaged Children and Their First School Experiences. ETS-Head Start Longitudinal Study: Preliminary Description of the Initial Sample Prior to School Enrollment. A Report in Two Volumes: Volume 1. Sponsored by the Office of Economic Opportunity. Princeton, NJ: Educational Testing Service, 1970. 300 pp. ED 047 797.

Silverstone, Naomi, editor. Programs for Infants and Young Children. Part I: Education and Day Care. Washington, D.C.: Appalachia Regional Commission, 1970. 221 pp. ED 047 807.

Sorensen, Philip H. Early Childhood Education Program Models for Indian Communities. Sponsored by OE/DHEW and Office of Indian Education. Oglala, SD: American Indian Resource Associates and Tsaile, AZ: Navajo Community College, 1974. 63 pp. ED 113 071.

Southeastern Education Laboratory. A Follow-Through Program Model. Rural Isolated Schools Program. Sponsored by OE/DHEW. Atlanta, GA: Southeastern Education Laboratory, 1969. 161 pp. ED 042 860.

Thomas, Susan B., compiler. Research on Approaches to Early Education: An Abstract Bibliography. Sponsored by the National Institute of Education/DHEW. Urbana, IL: ERIC Clearinghouse on Early Childhood Education, 1974. 84 pp. ED 092 262.

Turnbull, B. "Research on Implementation: An Annotated Bibliography." Cambridge, MA: Huron Institute, 1976. 31 pp.

Valencia, Atilana A. Identification and Assessment of Ongoing Educational and Community Programs for Spanish Speaking People. A Report Submitted to the Southwest Council of La Raza, Phoenix, Arizona. Albuquerque, NM: Southwestern Cooperative Educational Laboratory, 1969. 116 pp. ED 028 013.

Villaume, John and Haney, Walt. Project Follow Through National Evaluation Bibliography. Cambridge, MA: Huron Institute. April 1977. 87 pp.

Weikart, David P. Letter to Walt Haney, Huron Institute, May 21, 1976. Ypsilanti, MI: High/Scope Educational Research Foundation.

White, Sheldon, Day, Mary Carol, Freeman, Phyllis K., Hantman, Stephen A., and Messenger, Katherine P. Federal Programs for Young Children: Review and Recommendations, Vol. II: Review of Evaluation Data for Federally Sponsored Projects for Children. Washington, D.C.: USGPO, 1973. ca. 400 pp.

Wisler, Carl E. Memorandum to Karl Hereford, Director, Program Planning and Evaluation, May 5, 1969. Washington, D.C.: Office of Program Planning and Evaluation.